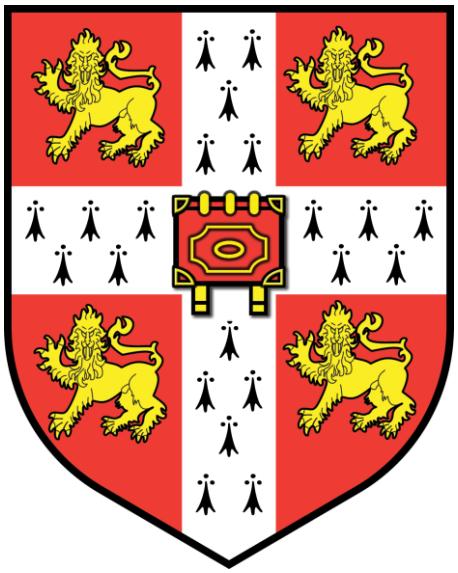
Biofunctionalised Coordination Cages and Quantified Speciation in Supramolecular Mixtures



Marion Kieffer

Darwin College

University of Cambridge

February 2019

This dissertation is submitted for the degree of Doctor of Philosophy.

Declaration

This dissertation, entitled *Biofunctionalised Coordination Cages and Quantified Speciation in Supramolecular Mixtures*, is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Projects contributions section and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed 60,000 words in length, the prescribed word limit for the Degree Committee for the Faculty of Physics and Chemistry.

Marion Kieffer

Biofunctionalised Coordination Cages and Quantified Speciation in Supramolecular Mixtures

Marion Kieffer

Summary

This thesis describes initial efforts towards the conjugation of biomolecules onto supramolecular self-assembled coordination cages in order to render the system biocompatible. Firstly, the compatibility and influence of coordination cages on a simple biomaterial was studied by the incorporation of tetrahedral cages in a peptide-based supramolecular gel. Secondly, bioconjugation of single amino acids and a tripeptide gelator onto coordination cages were investigated, showing that the properties of the peptide such as chirality and gelation were transferred to the whole system. Finally, a water-soluble cage was synthesised using *N*-acetylatedgalactoamine building blocks to allow ligation of the cage onto the asialoglycoprotein receptor of hepatocytes.

Deciphering the self-assembly rules within complex self-assembled libraries involves new characterisation challenges. In a second section of this thesis the potential to assess individual outcomes of self-sorting experiments by mass spectrometric techniques was investigated in three case-studies. The relative energies of heteroleptic structures compared to the more stable homoleptic as well as the effect of anion binding on a dynamic library of self-assembled tetrahedra was investigated. Furthermore, quantitative information on speciation within mixtures was obtained for a complex system of self-assembled scalenohedra and pseudo-octahedra.

Le vrai point d'honneur n'est pas d'être toujours dans le vrai. Il est d'oser, de proposer des idées neuves, et ensuite de les vérifier.

The real point is not to always be right. It is to dare, to propose new ideas, and then to validate them.

PIERRE-GILES DE GENNES

from *La Science des Rêves*, 2007

Abstract

Coordination cages formed by supramolecular self-assembly are excellent candidates for the selective encapsulation of molecules and sheltering these guests from their environment after uptake. The easily tuneable nature of these 3D-structures can make them versatile carriers in applications such as drug delivery. However, to achieve this, the capsules need to be converted into biocompatible vehicles by the conjugation of biomolecules onto the cages' surface for example. This thesis describes initial efforts towards this goal. Firstly, the compatibility and influence of coordination cages on a simple biomaterial was studied by the incorporation of tetrahedral cages in a peptide-based supramolecular gel. Rheological changes of the macroscopic material were observed while the host-guest properties of the cages remained unchanged, enabling the chemical segregation of guests. Secondly, bioconjugation of single amino acids and a tripeptide gelator onto coordination cages were investigated, showing that the properties of the peptide such as chirality and gelation were transferred to the whole system. Addition of a photoacid generator to the organo-gel formed triggered the reversible gel-sol transition of the material under alternating cycles of light irradiation and darkness. Finally, a water-soluble cage was synthesised using *N*-acetylatedgalactoamine building blocks. The geometry and organisation of the sugar biotags were engineered to allow ligation of the cage onto the asialoglycoprotein receptor of hepatocytes with the goal of realising the cellular internalisation of the complex and its cargo.

In synthetic supramolecular systems, molecules can interact in different ways to yield complex self-assembled libraries. Multiple components combined can result in the integrative formation of single products or, most often, multiple self-assembled products. Deciphering the self-assembly rules within such systems involves new characterisation challenges: dynamic, low-symmetry products are difficult to detect and identify by NMR spectroscopy and labile species can re-equilibrate after chromatographic separation. Hence, in this thesis the potential to assess individual outcomes of self-sorting experiments by mass spectrometric techniques was investigated in three case-studies. A new methodology to calculate the relative energies of heteroleptic structures compared to the more stable homoleptic was developed, allowing for the quantification of each ligand's structural preferences. Following a similar approach, the effect of anion binding on a dynamic library of self-assembled tetrahedra was probed. Finally, quantitative information on speciation within mixtures was obtained for a complex system of self-assembled scalenohedra and pseudo-octahedra.

Publications and copyright information

Parts of this dissertation have been published in the following peer-reviewed journals:

[1] M. Kieffer, B. S. Pilgrim, T. K. Ronson, D. A. Roberts, M. Aleksanyan, J. R. Nitschke, Perfluorinated Ligands Induce Meridional Metal Stereochemistry to Generate M_8L_{12}, $M_{10}L_{15}$, and $M_{12}L_{18}$ Prisms. *J. Am. Chem. Soc.* **2016**, *138*, 6813-6821.

[2] F. J. Rizzuto, M. Kieffer, J. R. Nitschke, Quantified Structural Speciation in Self-sorted $Co^{II}_6L_4$ Cage Systems. *Chem. Sci.* **2018**, *9*, 1925-1930.

[3] M. Kieffer*, A. García Fernández*, C. J. E. Haynes, S. Kralj, D. Iglesias, J. R. Nitschke, S. Marchesan, Embedding and Positioning of Two $Fe^{II}_4L_4$ Cages in Supramolecular Tripeptide Gels for Selective Chemical Segregation. *Angew. Chem. Int. Ed.* **2019,** doi:10.1002/anie.201900429.

Any work included in this dissertation that has been published has been reproduced herein with permission.

Parts of this dissertation are in preparation for publication:

[4] M. Kieffer, R. Bilbeisi, J. Thoburn, J. Clegg, J. R. Nitschke, Species amplification in a library of $Co^{II}_4L_4$ cages. *Manuscript in preparation.*

Additional collaborative work was carried during the course of this PhD project, but not included in this dissertation:

[5] E. G. Percástegui, J. Mosquera, T. K. Ronson, A. J. Plajer, M. Kieffer, J. R. Nitschke, Waterproof architectures through subcomponent self-assembly. *Chem. Sci.* **2019**, *10*, 2006-2018.

[6] J. Zhu, C. J. E. Haynes, M. Kieffer, J. L. Greenfield, R. D. Greenhalgh, J. R. Nitschke, U. F. Keyser, $Fe^{II}_4L_4$ Tetrahedra Detect Base Mismatch in DNA by Fluorescence Quenching. *Submitted*.

* denotes joint first author

Projects contributions

Parts of the work in this dissertation were undertaken in collaboration with scientists both at the University of Cambridge and outside. All contributions are explained below, as well as in each chapter.

The work presented in Chapter 3 was carried out in collaboration with the Marchesan Group at the University of Trieste. Tripeptide **3.C** was designed and synthesised by Dr. Ana García Fernández and part of the material characterisation (TEM, EDX and rheology) were carried out by Dr. Slavko Kralj and Dr. Ana García Fernández.

The work presented in Chapter 4 was carried out with the help of Mr. Shihao Zang (a visiting BSc student from Peking University). Mr. Shihao Zang synthesized the photo-acid generator **MEH** and carried out preliminary investigations on the compatibility of this compound with $Fe^{II}_4L_4$ cages and on the reversible sol-gel transition under the guidance of Marion Kieffer. Dr. Cally Haynes contributed to the synthesis of anilines **4.A** – **4.E**. CD measurements were performed by Mr. Jake Greenfield. As in Chapter 3, tripeptide **3.C** was synthesized by Dr. Ana García Fernández.

The design of the project described in Chapter 5 was undertaken as a collaboration between Dr. Carsten Behrens, Dr. Thomas Nielsen (Novo Nordisk, Denmark) and Marion Kieffer. Fluorescence spectra were acquired by Mr. Jake Greenfield. Cell studies are being performed by scientists in Novo Nordisk. Triamine **5.N** was synthesized by Dr. Dawei Zhang.

The work presented in Chapter 6 was undertaken as a collaboration with previous and current members of the Nitschke group at the University of Cambridge. Dr. Ben Pilgrim contributed to initial discussions on the development of the quantitative mass spectrometry analysis in section 6.2.2. Dr. Felix Rizzuto designed and characterised the chemical system used in section 6.3. Dr. Rana Bilbeisi set precedent for the results discussed in section 6.4 and grew the crystal of cages **6.9** and **6.10** and Dr. Jack Clegg solved the crystal structures of these cages. Finally Prof. John Thoburn created the Mathematica program and preformed the curve fitting to obtain the binding affinities of guests with **6.10**.

Note: A preliminary version of the results presented in section 6.2 of chapter 6 have been previously submitted by Marion Kieffer as part of a dissertation for the Master of Philosophy degree of the University of Cambridge.

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Jonathan Nitschke for giving me the opportunity to stay for a PhD after I finished my MPhil in his group. I am very grateful for your guidance and support, and most of all, for the freedom you gave me to come up with and explore my own ideas throughout the years. More than once, you have encouraged me to investigate things far beyond the point where I would have usually given up.

To all my collaborators in the Nitschke group, thanks for all the things you taught me: you have made me a better researcher and scientist than I could have hoped for. Special mentions go to Cally for always being a chair spin away to answer any questions, more or less relevant to science, and to Felix for "sanity-checking" my ideas and always answering honestly my sometime stupid questions. Nothing would have been the same without all the members of Lab 354/353, who made every day brighter than the gloomy British weather allowed for; keep doing so! Thanks Lillian for keeping the bay playlist spot on, Larissa for the endless discussions about MS, Tanya for always being a reliable source of knowledge, Roy for being my to-go IT guy, Hugh for keeping my caffeine levels up and Nguyen for keeping my sugar levels up.

Outside the lab, I must also thank Dr. Thomas Nielsen and Dr.Carsten Beherens, as well as all the members of the Protein and Peptide Chemistry team in Novo Nordisk for welcoming me and teaching me about the marvels of peptide synthesis. Most of the work done in this thesis would not have been possible without the help of Mr. Duncan Howe, Dr. Peter Grice, and Mr. Andrew Mason from the Cambridge NMR service. You have all taught me so much about NMR and have always been there when I needed help.

I would also like to thank my parents for supporting me for whole of my studies, even if they have no clue what I am doing. Jan, thank you for putting up with me, enduring my incessant ranting and making sure I never ran out of orange juice.

Finally, I would not have retained half of my sanity without the support of the amazing friends I made at Darwin College and through its Boat Club. If I have to remember only one of all the things I shared with my favourite robust ladies, it would be that everything, in science and life in general, seems much easier when you compare it to a 2K test.

On a more formal note, I would like to thank the European Union's Horizon 2020 Marie Sklodowska-Curie Actions for funding.

Abbreviations

2D	two-dimensional
3D	three-dimensional
Ala	alanine
Arg	arginine
Asp	aspartic acid
CD	Circular Dichroism
COSY	Correlation Spectroscopy
DIPEA	*N*,*N*-diisopropylethylamine
DOSY	Diffusion Ordered Spectroscopy
EDX	Energy Dispersive X-ray
eq.	equivalents
ESI	Electrospray Ionisation
fac	facial
FESEM	Field Emission Scanning Electron Microscope
Fmoc	fluorenylmethyloxycarbonyl protecting group
GalNAc	N-acetylgalactosamine
gel-sol	gel to solution
Gly	glycine
HBTU	O-(benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
HOAt	1-hydroxy-7-azabenzotriazole
HPLC	High Performance Liquid Chromatography
HR	High Resolution
HSQC	Heteronuclear Signle Quantum Coherence spectroscopy
IM	Ion Mobility
Leu	leucine
LMWG	Low Molecular Weight Gelator
LR	Low Resolution
mer	meridional
MLCT	Metal-Ligand Charge Transfer
MOF	Metal-Organic Framework
MS	Mass Spectrometry
M_xL_y	metal complex with *x* metals (M) and *y* ligands (L)
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect correlation Spectroscopy
NTf_2^-	triflimide anion; bis(trifluoromethanesulfonyl)imide
OTf^-	triflate anion; trifluoromethanesulfonate
PAG	Photo Acid Generator
PBS	Phosphate Buffer Saline
Phe	phenylalanine
rt	room temperature
sol-gel	solution to gel

STD	Standard Deviation
TBA^+	tetrabutyl ammonium cation
TEM	Transmission Electron Microscopy
TFA^-	trifluoroacetate anion
TLC	Thin Layer Chromatography
Trp	trypotophan
UV-vis	ultraviolet-visible

Table of contents

Chapter 1

Introduction

1.1 Understanding self-assembled coordination cages

Supramolecular chemistry focuses on the synergistic combination of discrete molecular subunits into systems *via* intermolecular interactions. Mixtures of compounds are led into defined products during self-assembly, driven by the formation of non-covalent interactions. Weaker than their covalent counterpart, H-bonds, π-π (and other electrostatic) interactions and metal-ligand coordinative bonds are key elements to achieve the reversibility necessary to form the thermodynamic product from a starting library.

1.1.1 General guidelines for design

Self-assembled coordination cages rely on the formation of dative bonds between metals and ligands, for the construction of 3D structures able to delimit the space between the outside or inside of the framework formed. The strong, yet labile, bonds formed along with the chelate effect, allow the system to rearrange into stable assemblies under thermodynamic control.

Combining this versatile type of interaction with specially designed organic ligands has led to a rapid increase in the number and type of architectures synthesized.[1-2] The interplay between the symmetry, the number of coordination sites and the size of the ligand along with the coordination geometry of the metal cation and the overall stoichiometry of the system influences the outcome of the reaction, yielding the most stable product. Highly tailorable assemblies are therefore formed following simple geometric rules of complementarity between the organic and inorganic molecular components of the system.

The reaction between the components of a chemical library will produce the most entropically and enthalpically stable architecture that satisfies both the coordination geometry of the metal and the coordinate vector of the organic molecule.[3] The ligands can only bind to the metal centre if their denticity and angle matches the metal coordination sphere.[4] The design of organic ligands and the metal chosen to occupy the edges and vertices are both paramount to obtain the desired shape (Figure 1.1).[5-6] The ligand can be mono or multi-dentate, singly or multi-branched, straight, bent or even concave or convex for multitopic molecules.

For example, Fujita *et al.* demonstrated the importance of the bend angle of a ditopic ligand in determining the final structure.[7] By iterative synthesis leading to ligands with subtly different curvatures, increasingly larger structures were self-assembled, from simple octahedron **1.1**,[8] to more complex cuboctahedron **1.2**,[9] rhombicuboctahedron **1.3**[10] and finally icosidodecahedron **1.4** (Figure 1.2).[11]

With these rules in hand, supramolecular chemists have been able to gain a better understanding in the self-assembly process, leading to rational design of ligands and subunits to obtain the desired structure, slowly moving away from serendipity.

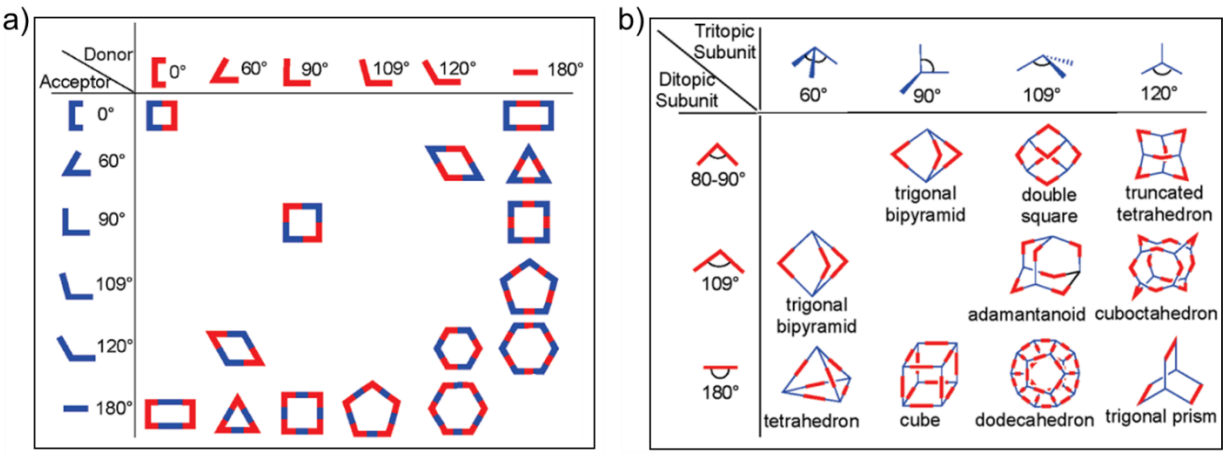


Figure 1.1 | Geometric complementarity of subunits in self-assembly of coordination cages. a) 2D convex polygons and canonical polyhedral are obtained from ditopic building blocks and b) 3D architectures are obtained from ditopic and tritopic blocks. Figures adapted with permission from reference 6.

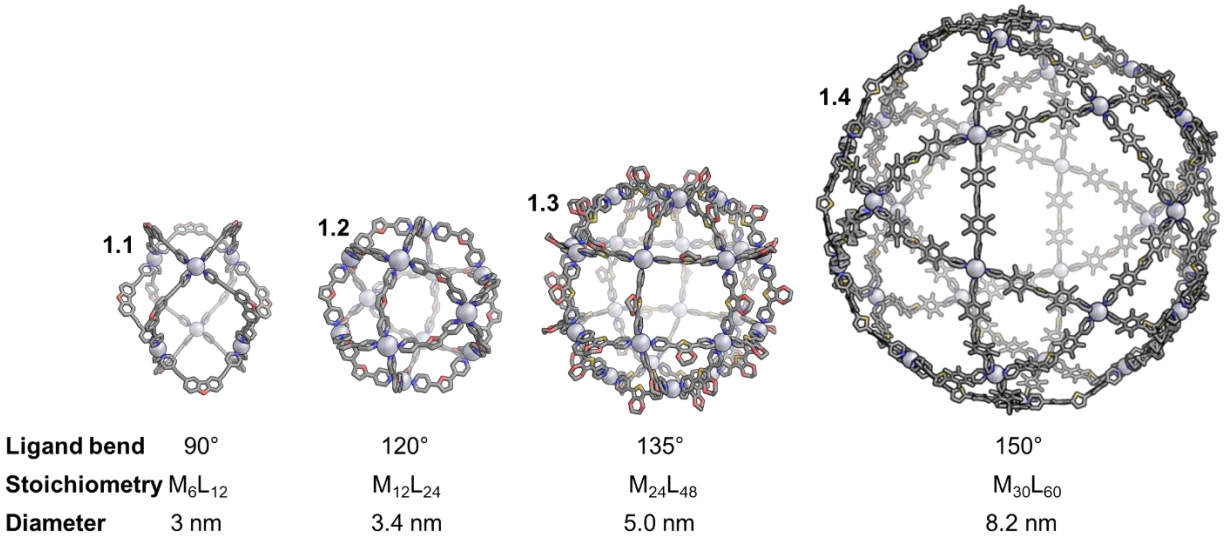


Figure 1.2 | Capsules of formulae M_xL_{2x} obtained by variation of the ligand bend from 90° to 150° leading to changes in the structure stoichiometry and therefore diameter.[8-11]

1.1.2 *Imine chemistry and subcomponent self-assembly*

Subcomponent self-assembly relies on the formation of both covalent bonds and metal-organic dative bonds through a metal-templation process. First investigated by Busch *et al.*,[12] this strategy was further developed by Hannon *et al.*[13] and later by Nitschke *et al.*[14] Used to stabilise copper mononuclear complexes,[14-15] the method has evolved to incorporate complex polydentate ligands[16-17] in a bottom-up approach, reacting separate organic subcomponents around a metal template instead of coordinating preformed ligands. Coordination leads to stabilisation of the covalent bond formed, but also of the metal, preventing its oxidation or reduction once coordinated.

Imine bonds may be reversibly cleaved and reformed, making them an ideal dynamic covalent bond to use in subcomponent self-assembly. As a product of the condensation of an aldehyde with an amine (mainly anilines and picolinaldehydes in this thesis), the imine bond formed is reversible through hydrolysis. However, once coordinated to a metal *via* its nitrogen donor atoms, the complex formed renders the imine bond stable, even in water in selected cases (Figure 1.3).[15]

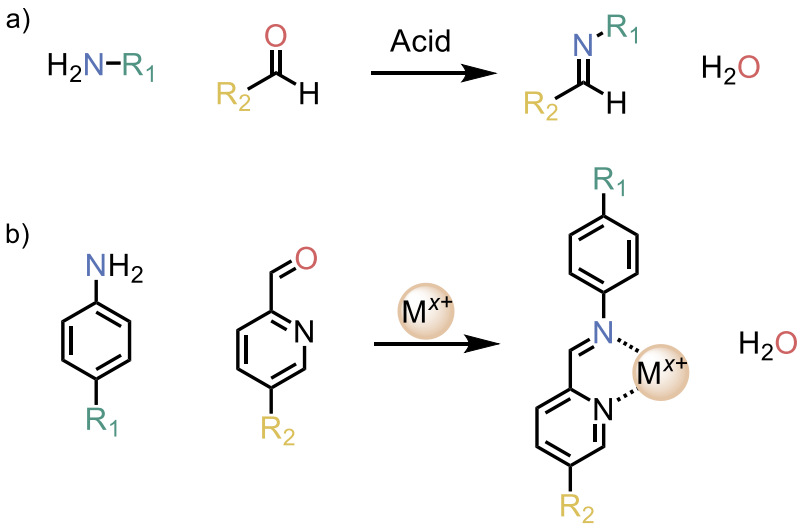


Figure 1.3 | Formation of an imine between an amine and an aldehyde. a) General case and b) specific case of the metal templated imine formation between a picolinaldehyde and an aniline.

This design allows building blocks to be readily switched, making it extremely versatile. The morphology and the resulting properties of a structure can be influenced by slight variations in the starting library. The structures yielded are thus varied and include helicates,[18] tetrahedra,[19] octahedra,[20] cubes,[21] prisms,[22] polymers[23] and many more.[24-25]

Furthermore, only one component of the library (pyridine, amine, metal or anion) needs to be changed to modify the behaviour of the whole system, limiting the synthetic effort required. For example, the amine used in the self-assembly process can undergo metathesis as another amine is added to solution.[26] This strategy has been applied by Nitschke *et al.* to generate a library of tetrahedra in a "cascade" manner by consecutive substitution of *p*-chloroaniline by *p*-toluidine and *p*-anisidine.[27] The Hammett parameter of the aniline was also shown to infuence the size of the architecture formed, going from smaller M_8L_{12} **1.5** to the larger $M_{10}L_{15}$ **1.6** or $M_{12}L_{18}$ prisms[28] or to influence the rate of reaction at a site on the coordination cage's ligand (Figure 1.4).[29]

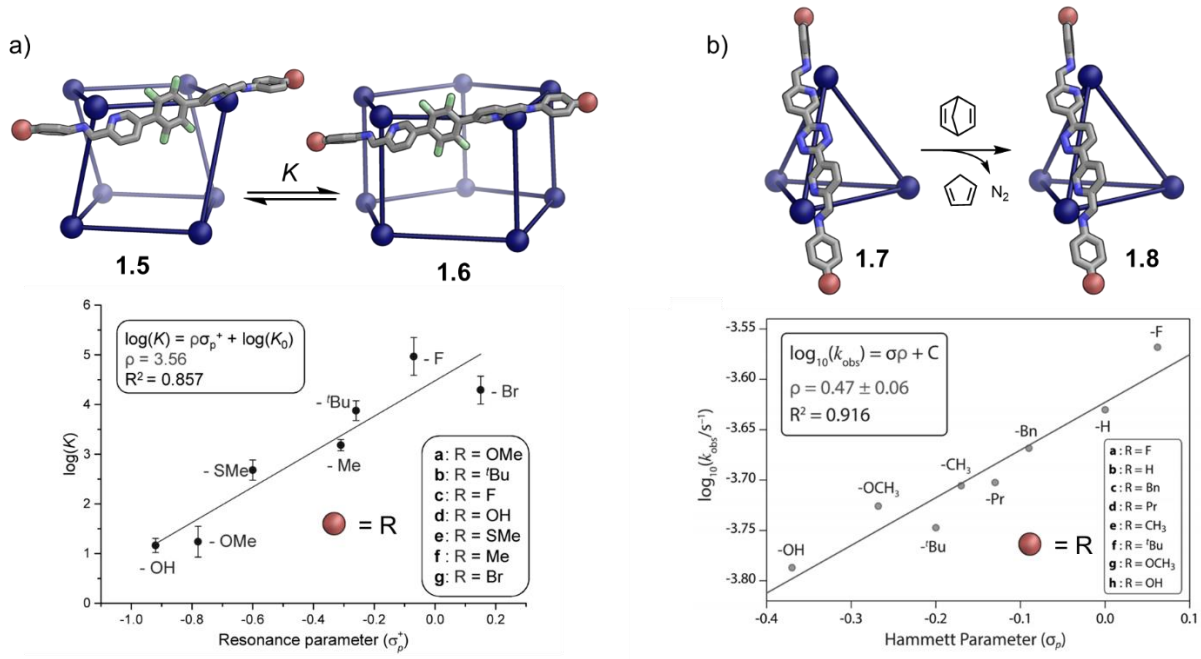


Figure 1.4 | Influence of the aniline's Hammett parameter on self-assembled cages displaying the pyridyl-imine coordination motif. a) The equilibrium between two structures **1.5** and **1.6** or b) the rate of conversion of **1.7** into **1.8** is affected, showing a linear free-energy relationship between log(*K*) and σ_p^+ or σ_p respectively.[28-29]

The power of subcomponent self-assembly can be illustrated by another example: the use of different metals and counter-ions. Recently, the role of the metals on the coordination cages stability in water has been unravelled in work by Nitschke *et al.* The authors showed that the half-life of structures could be increased from minutes to months by changing the metal from Cd^{II} to Fe^{II} or Ni^{II}.[30] The role of the counter-anion was also deciphered, where the hydrophobicity of the anion was shown to be paramount for the solubility of the coordination cage in different phases, bringing water solubility[31] or phase transfer ability to cages.[32]

1.1.3 Geometry and chirality

Transition metals are extensively used in supramolecular chemistry due to their fixed, well-defined coordination geometries, making them good templates for self-assembly and rational design. The metal d-orbital occupancy dictates the lability of the metal and the geometry of the coordination sphere. Most commonly tetrahedral, square planar and octahedral metals are chosen for the formation of coordination cages.[6] For octahedral coordination which is our focus in this thesis, either six monodentate, three bidentate or two tridentate ligands bind to the metal centre *via* dative interactions.

Bidentate ligands, if asymmetric, can bind in two different fashions, and will, if no other parameters influence the binding, yield a statistical mixture of *meridional* and *facial* geometries in a 3:1 ratio. The *facial* (*fac*) coordination is defined by the geometry in which the three imine nitrogen atoms define one triangular face of the octahedral coordination sphere, as opposed to the *meridional* (*mer*) coordination where the three imine nitrogen atoms define a plane that includes the metal centre (Figure 1.5 a).[33] The formation of *fac* or *mer* vertices therefore has a strong influence on the resulting structure. The "wider" opening in the angle between ligands in the case of *mer* vertices as opposed to *fac* allows for larger structures with higher nuclearity to form like prisms[22, 34] or icosahedra[25] for example.

In addition two different optical isomers exist (Δ and Λ) contributing to the complexity of structures that can be achieved using the same ligand and transition metal (Figure 1.5 b). Usually both isomers are present in solution if no influencing factors are present. Otherwise, when one optical isomer is present in larger proportions than the other, chiral induction is observed which can be partial or complete. In the latter case, enantiopure cages are obtained, most often resulting from the use of chiral ligands in self-assembly.[35-37]

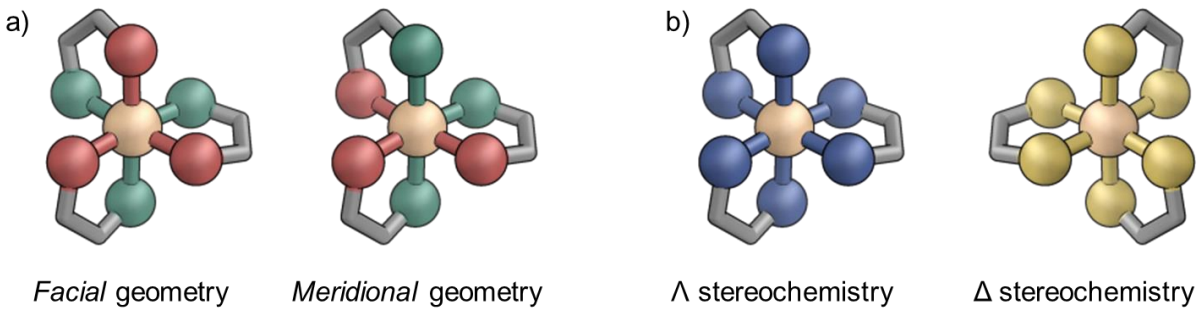


Figure 1.5 | a) *fac* and *mer* geometries and b) Λ and Δ stereochemistry obtained for an octahedral metal centre when reacted with three bidentate ligands.

Coordination cages can also present both optical isomers of the metal vertices in the same structure. For example, for tetrahedral cages with edge-bridging ligands (where the organic component of the coordination cage lies along the edge of the polyhedron formed, as opposed to face-capped cages where the organic component closes the faces of the structure), three different isomers can exist. When the same optical isomer is present at all four vertices (ΔΔΔΔ or ΛΛΛΛ), the overall structure displays *T*-symmetry. When one vertex presents the opposite isomer (ΔΔΔΛ or ΛΛΛΔ) the structure has C_3-symmetry. Finally when two vertices are Δ and the other two are Λ the structure has S_4-symmetry (Figure 1.6 a).[38] However, most edge-bridged tetrahedra reported form the *T*-symmetric isomer only.

Face-capped tetrahedra are also usually *T*-symmetric, due to the propeller shape of their tritopic ligands enforcing homochirality between all metal vertices. The overall symmetry of the structure can nonetheless be changed by employing either ligands that can be present in 2 conformations (*endo* and *exo* for example) or by employing two different types of ligand. Similarly to edge-bridged tetrahedra, the coordination cages obtained can be defined as *T*-symmetric, C_3-symmetric or C_2-symmetric (Figure 1.6 b).

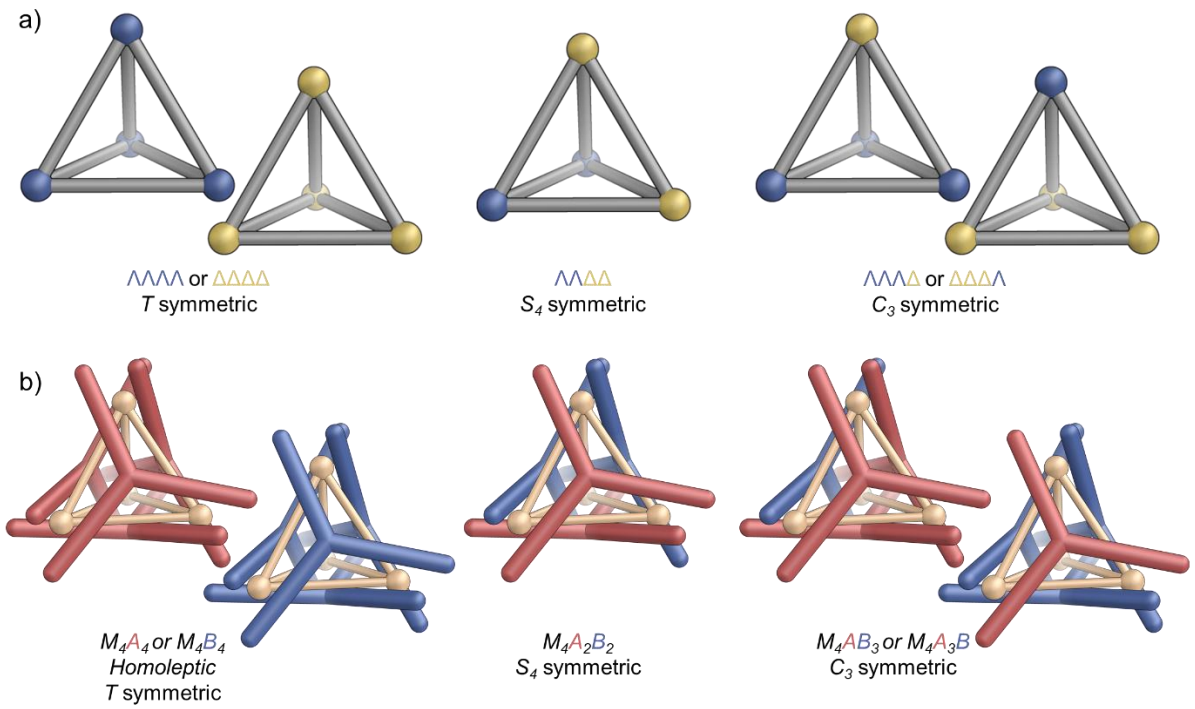


Figure 1.6 | a) Different optical isomers for edge-bridged tetrahedra, the blue and yellow balls representing Λ and Δ vertices respectively. b) Face-capped tetrahedra of different symmetry based on the type of ligand employed.

1.1.4 Complexity in systems: mixed ligands and self-sorting

The complexity of biological macromolecular systems arise from the assembly of multiple molecular components arranged in a careful and controlled way. The variety of functionalities achieved by these systems such as catalysis or molecular recognition, to cite only a few, are tailored by the properties of the components used. Similarly, in coordination-driven self-assembly, complexity and an increase in functionality of the 3D structures formed can be obtained by using a combination of organic building blocks.

When multiple components are used in self-assembly, a library is formed. Self-sorting occurs in the system if there is a degree of recognition between the molecules in the library. If the building blocks have an affinity for themselves only, the process is referred to as narcissistic self-sorting, whereas social self-sorting occurs if building blocks of a different nature get incorporated into the final products (Figure 1.7 a-b).[39] Cooperation between the building blocks can be complete, in which case a single product is formed, referred to as integrative sorting (Figure 1.7 c). In the cases where the socialisation is not complete, multiple products are formed during non-integrative processes.[40]

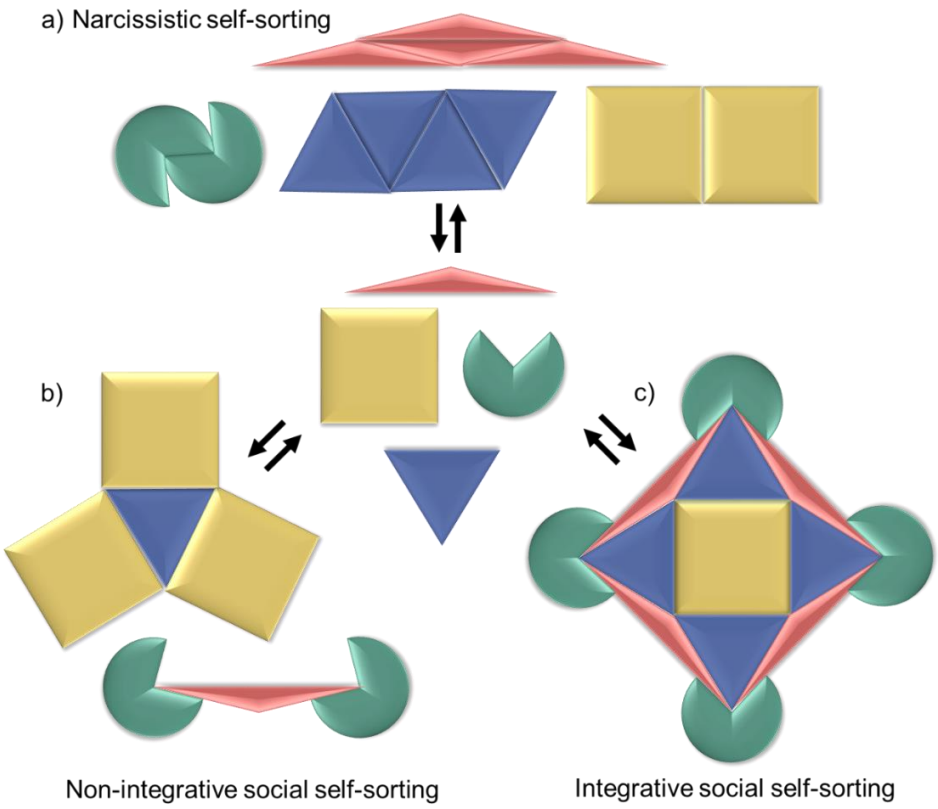


Figure 1.7 | Scheme of the different self-sorting regimes possible, where components of the library sort a) narcissistically, b) socially and c) integratively.

Because heteroleptic systems (made of different ligands, as opposed to homoleptic) are built through self-sorting, new properties emerge from these,[41] one of which can be the formation of new, unexpected architectures.

For example, Fujita *et al.* employed a combination of two similarly shaped ligands, but varying in size, to generate a $Pd_{12}L_{12}L'_{12}$ pseudocantellated tetrahedra **1.9** (Figure 1.8). The system underwent integrative self-sorting, with a single highly desymmetrised product formed.[42] Another example is the formation of a triple-decker $Zn^{II}_4L_4L'_2$ sandwich **1.10** when both NDI- and pyrene-based ligands were employed, as described by Nitschke *et al.* (Figure 1.8).[43]

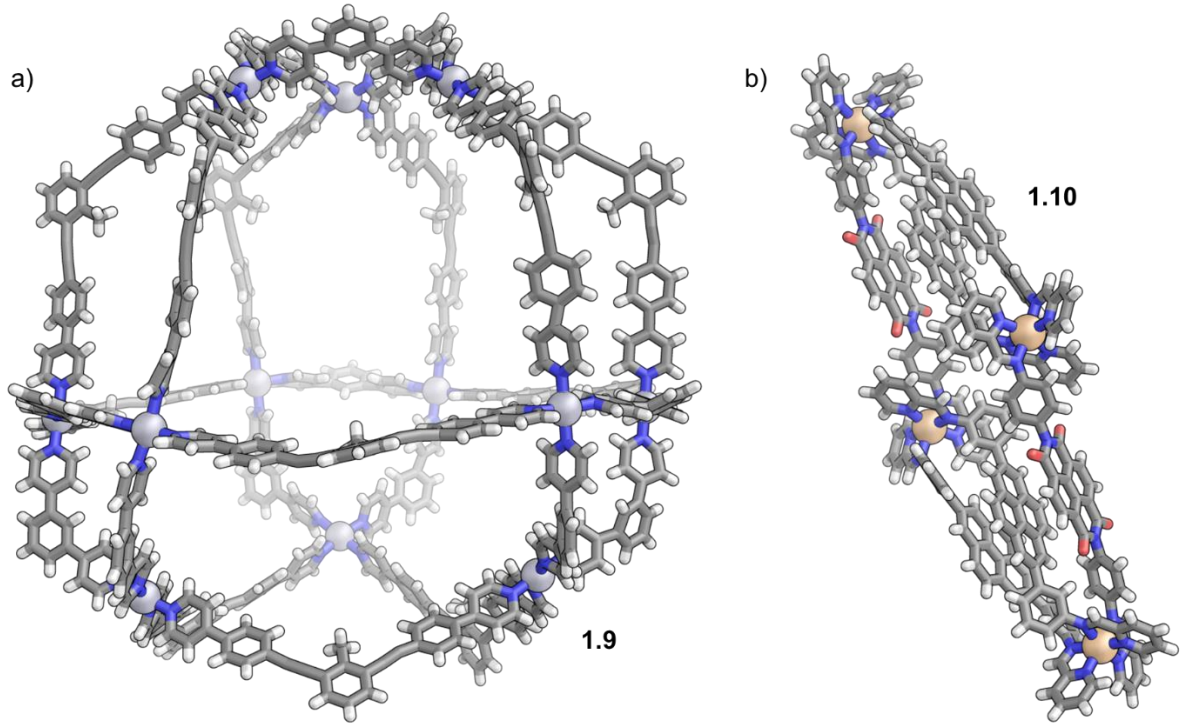


Figure 1.8 | Crystal structures of heteroleptic coordination structures formed from two different organic ligands. a) $Pd_{12}L_{12}L'_{12}$ pseudocantellated tetrahedra **1.9** and b) triple-decker $Zn^{II}_4L_4L'_2$ sandwich complex **1.10**.[42-43]

Incorporating a variety of ligands carrying more than a single functionality can confer new properties to the heteroleptic cages formed, such as catalysis[44] or new host guest properties.[45] A well-known example is the formation of stable mono- and di-nucleotide duplexes in the cavity of self-assembled heteroleptic prisms as demonstrated by Fujita *et al.*[46] Recent unpublished work from the Nitschke group showed the formation of a triangular prism from porphyrin-based tetratopic and triazine-based tritopic ligands presenting both the right cavity shape as well as electronic interior to encapsulate a series of steroids.[47]

1.2 Host-guest chemistry in coordination cages

One of the features of molecular cages or containers, as their name indicates, is the presence of a well-defined internal void capable of accommodating guests. The conditions for guest binding within the cavity of cages are diverse and the rules governing the specificity of molecular interactions between host and guest are complex.[48] However, thanks to their high tuneability, coordination cages are good candidates for the study of host-guest recognition. With a few rules in hand, it is possible to engineer molecular recognition by rationally designing the appropriate cage.

1.2.1 General rules for encapsulation

The thermodynamic process of guest encapsulation within the cavity of a cage relies on both entropic and enthalpic contributions. Most often, the major entropic driving force results from the solvent being both expelled from the cavity and desolvated from the guest, whereas enthalpic gains are generated by positive cooperative interactions. The predominant factor for successful encapsulation lies in the size and shape match between the host and the guest. It was shown that optimal binding could be achieved when approximately 55% of the available cavity is occupied, leading to the so called "55% rule".[49] This trend was investigated experimentally by Hunter, Ward *et al.*, in a coordination cage giving an experimental value of *ca.* 50% for cavity occupancy before the binding strength started to decrease.[50]

Quite often the cavity of coordination cages presents a different environment from the bulk solution, such that complementary interactions increase the binding affinity. These strong intermolecular forces include hydrophobic interactions,[51] hydrogen bonding[52] and π-stacking[53] amongst others. The ability of a cage to adapt its cavity to match a guest is a good example. Velders, Severin *et al.* reported cage **1.11**, able to expand its cavity in order to bind two coronene molecules, thus increasing intermolecular interactions (Figure 1.9).[54] Closing off the faces of the cage is another common strategy to increase enthalpically-favourable interactions. Nitschke *et al.* elegantly demonstrated that increasing the size of the aromatic panel situated on $Fe^{II}_4L_6$ tetrahedra **1.12-X** was directly related to changes in guest binding strength, with more enclosed cages encapsulating a larger range of guests (Figure 1.10).[55]

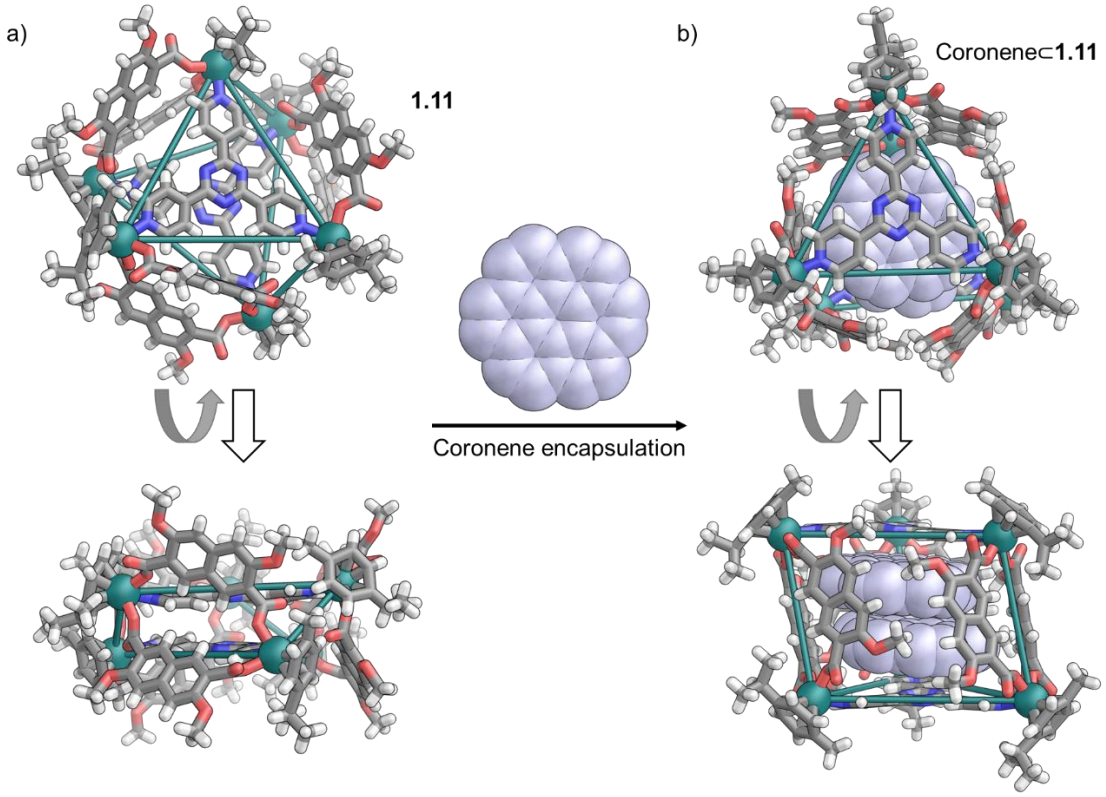
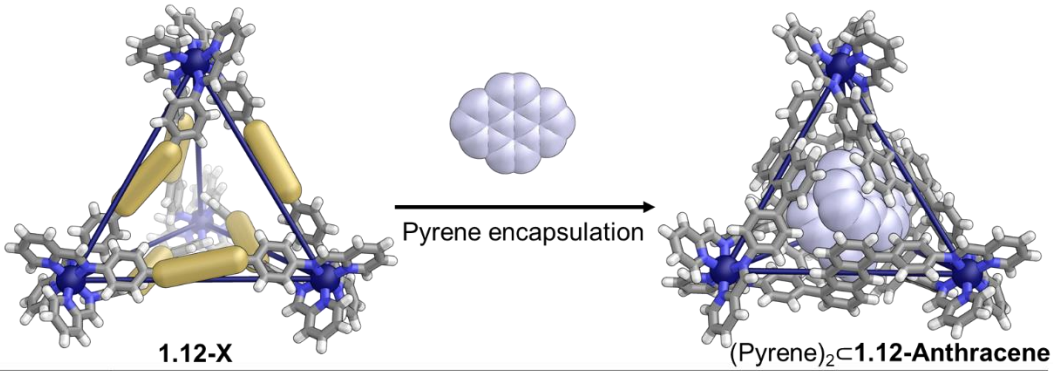


Figure 1.9 | Crystal structures of $Ru^{II}_6L_2L'_6$ **1.11** before and after encapsulation of two coronene (purple) which resulted in a conformational switch from a collapsed to an extended trigonal prismatic structure.[54]



Guest	Encapsulation observed (NMR rate of exchange at 298 K)								
≡									
C_{60}	Yes (slow)	No	No	No	Yes (slow)	No	No	Yes (slow)	No
C_{70}	Yes (slow)	No	Yes (slow)	No	Yes (slow)	No	No	Yes (slow)	No
coronene	Yes (slow)	Yes (fast)	No	No	No	No	No	Yes (slow)	No
perylene	Yes (fast)	Yes (fast)	Yes (fast)	No	No	No	No	Yes (fast)	No
pyrene	Yes (fast)	No	No	Yes (fast)	Yes (slow)	Yes (fast)	No	Yes (fast)	No
triphenylene	Yes (fast)	No	No	No	Yes (fast)	Yes (fast)	No	Yes (fast)	No
anthracene	Yes (fast)	No	No	No	No	Yes (fast)	No	Yes (fast)	No
cholesterol	Yes (fast)	Yes (fast)	No	No	Yes (fast)	No	No	Yes (fast)	No

Figure 1.10 | Scheme for tetrahedron **1.12-X** with X representing the different panels on the centre of the ligand (yellow stick) and crystal structure of **1.12-Anthracene** after encapsulation of two pyrene molecule. Table showing the relation between the binding of different guests and the type of aromatic panel with more enclosed cages leading to stronger binding.[55] Figures adapted with permission from reference 55.

Finally, charge complementarity between coordination cages (often positively charged) and guests (often negatively charged) must also be accounted for to assess guest binding.[56-57] Poor size-match between host and guest can be overcome thanks to Coulombic attraction, with the caveat of overcoming electrostatic interactions such as ion pairing between the anionic guest and its counter-cation.

1.2.2 Templates, an essential type of guest

Binding of guests, mainly anionic in cationic cages, is commonly used to direct the formation of architectures. The templation process can lead to the formation of unexpected, new geometries instead of the otherwise obtained discrete structure without a template. More importantly, the use of a template in self-assembly has been shown to sometimes be necessary to direct the subcomponents from a dynamic library to the formation of a single or multiple discrete species.

Nitschke *et al.* have extensively exploited this strategy to induce the reconfiguration of cages upon addition of a template[58] or to drive structural adaptation[59] for example. In one case different anions were used to form a series of discrete architectures from a dynamic library obtained when NTf_2^- was employed (Figure 1.11).[60] Upon binding of OTf^-, tetrahedron **1.13** and grid **1.14** were formed, whereas BF_4^- drove the mixture towards tetrahedron **1.13** and grid **1.14** as the kinetic product which then reconverted into pentagonal prism **1.15,** inferred to be the thermodynamic product. When ClO_4^- or NO_3^- were used, single products were formed, respectively pentagonal prism **1.15** or cuboid **1.16**. Similarly, Ward *et al.* used ClO_4^- and BF_4^- to drive the formation of $Co^{II}_4L_6$ tetrahedral capsules as opposed to the $Co^{II}_2L_3$ helicate previously obtained.[61]

Ureas can strongly coordinate to anions and, as such, have been incorporated in a multitude of ligands used in self-assembly. For example, Hay *et al.* have used this motif to generate M_4L_6 cages forming exclusively around oxoanions EO_4^{n-} (E = S, Se, Cr, Mo, W, P) by binding of four urea-based ligands around a single anion.[62]

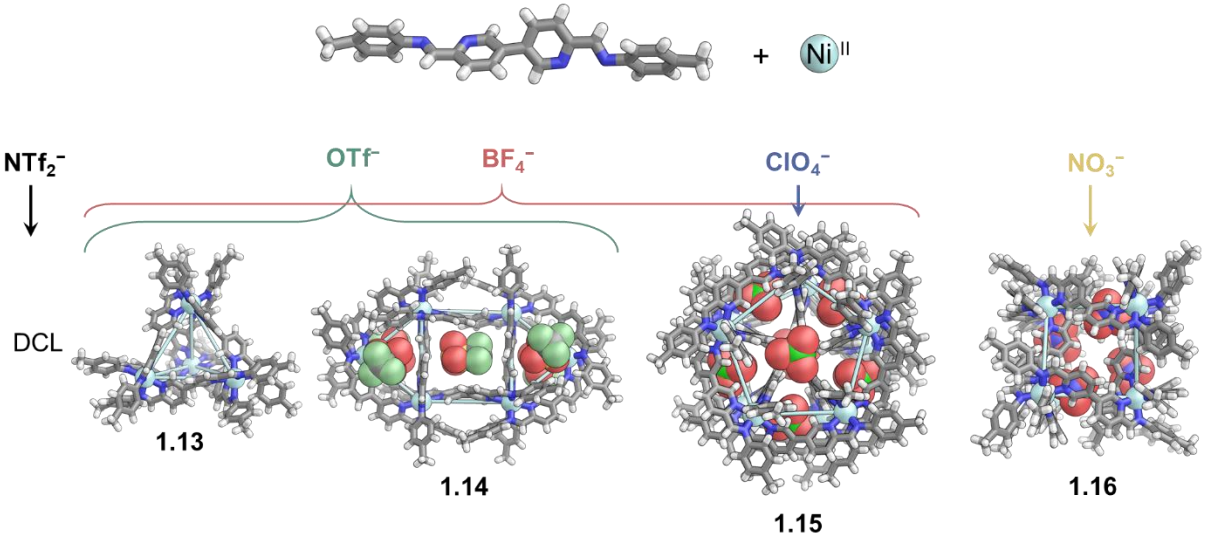


Figure 1.11 | A series of four discrete $Ni^{II}_xL_y$ supramolecular coordination architectures were obtained from the same ligand and metal, by varying the anion used as a template. NTf_2^- gave a non-discrete dynamic library, OTf^- gave tetrahedron **1.13** and grid **1.14**, BF_4^- gave tetrahedron **1.13** and grid **1.14** as the kinetic product which then reconverted into the thermodynamically stable pentagonal prism **1.15**, ClO_4^- gave pentagonal prism **1.15** solely when traces of Cl^- were present and NO_3^- gave cuboid **1.16** and precipitate.[60]

1.2.3 Chemical separation

Separation of molecular compounds is a challenge which chemists have been working on for decades, developing techniques ranging from simple extractions to more sophisticated approaches such as chromatography. More recently, the use of molecular hosts able to bind compounds of interest has emerged as a new strategy to separate mixtures of molecules or to recover precious compounds. Small organic receptors able to strongly bind anions have been extensively employed to extract harmful anions from the environment (TcO_4^-, CN^- or AsO_4^- amongst others)[63] or to recover gold[64] for example. Metal-organic frameworks (MOFs) have also shown great potential for chemical separation of a wide range of molecules (gases, anions, enantiomers, structural isomers etc).[65]

Similarly to MOFs, one could imagine that the high selectivity of a discrete coordination cage for a certain class of compound could be used towards chemical separation. However their use in this context is still embryonic. One of the main challenges to overcome in order to enable the use of coordination cages as chemical separation tools is the extraction of the cage from the bulk solution and the subsequent recovery of the compound of interest. To do so, phase transfer has been explored as a strategy.

Nitschke *et al.* demonstrated that changing the anion associated with a cationic cage allows, based on their hydrophobicity, the selective transfer of one cage over another between phases, thus transporting and separating their encapsulated molecular cargoes.[32] The chemical separation and recovery of ReO_4^- from other anions was achieved via this method.[66] $Fe^{II}_4L_4$ molecular cage **1.17** capable of anion recognition in water was employed.[59] When dissolved in CH_3NO_2, **1.17** was shown to selectively encapsulate ReO_4^-. After removing the water layer and evaporating the organic solvent, transfer to EtOAc led to the decomposition of **1.17** over time, thus releasing the guest in a clean water phase. Finally, evaporation of the EtOAc and suspension of the subcomponents from **1.17** in CH_3CN allowed the reassembly and recovery of **1.17** (Figure 1.12).

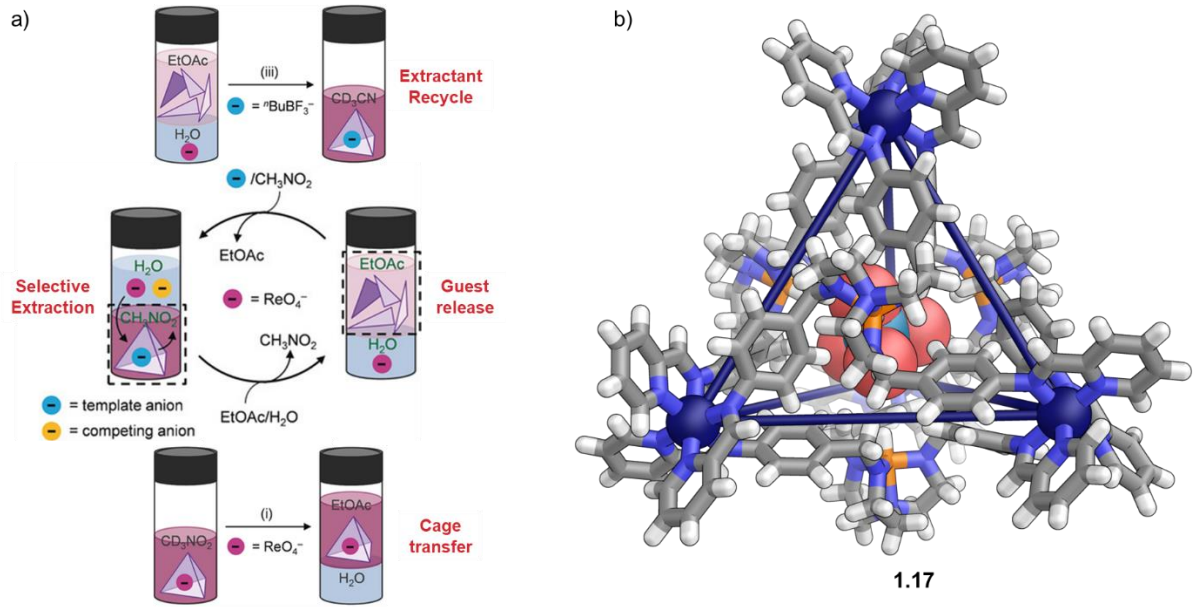
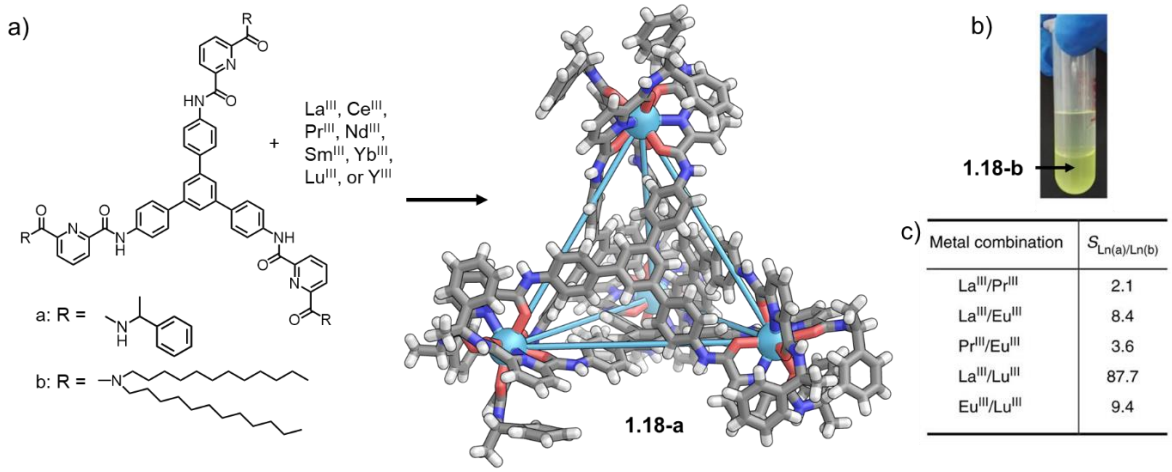


Figure 1.12 | a) Scheme depicting the cyclic extraction and recovery of ReO_4^- from a mixture of anions in water by **1.17**. b) Crystal structure of $ReO_4^- \subset$ **1.17**.[66] Figures adapted with permission from reference 66.

Instead of employing the cage cavity as a means to separate molecules, Sun *et al.* recently exploited self-assembly to selectively coordinate specific ions from a mixture of lanthanides.[67] A pyridyl-diamide ligand was shown to self-assemble with a range of lanthanide metal ions (La^{III}, Ce^{III}, Pr^{III}, Nd^{III}, Sm^{III}, Yb^{III}, Lu^{III}, or Y^{III}) to give cage **1.18-a**, with higher affinity for some of these ions. Addition of alkyl chains at the periphery of the ligand allowed transfer of the cage formed from water to chloroform. Assembling cage **1.18-b** in the presence of a mixture of La^{III} cations and partitioning the mixture between chloroform and water resulted in the extraction of the metal preferentially bound, which could then be recovered by disassembling cage **1.18-b** in CH_3CN (Figure 1.13).



Metal combination	$S_{Ln(a)/Ln(b)}$
La^{III}/Pr^{III}	2.1
La^{III}/Eu^{III}	8.4
Pr^{III}/Eu^{III}	3.6
La^{III}/Lu^{III}	87.7
Eu^{III}/Lu^{III}	9.4

Figure 1.13 | a) Self-assembly of pyridyl-diamide ligands bearing two different groups (a or b) and a series of lanthanide metal ions gave tetrahedral cage **1.18**. The crystal structure of cage **1.18-a** is given here. b) **1.18-b** could be extracted in $CHCl_3$ while the uncomplexed lanthanide cations remained in the water phase. c) Separation factors obtained from lanthanide solvent extraction.[67] Figures adapted with permission from reference 67.

1.2.4 Applications to drug delivery

Currently, medical treatments rely mostly on small drug molecules administered orally or injected directly into in the patient's bloodstream, often leading to severe toxicity or the development of multi-drug resistance. The use of nanocarriers building on the enhanced permeability and retention effect has been explored to improve drug efficiency.[68] However, poor tissue penetration of nanomedicines and lack of selectivity have been highlighted as the major issues encountered, resulting from poor solubility, low stability of the drugs and carriers, or conflicting charge interactions.[69]

Coordination cages have recently been investigated as candidates for drug delivery, based on their potential to protect the drug in the body. The molecular recognition and unique environment found inside the coordination cage can lead to selective encapsulation of biologically relevant molecules. For example, Yoshizawa *et al.* have demonstrated the ability of $Pd^{II}_2L_4$ capsule **1.19** to encapsulate lactic acid derivatives[70] while Fujita *et al.* have shown the sequence selective uptake of peptides into $Pd^{II}_6L_4$ octahedral cage **1.21**[71] (Figure 1.14 a and c).

Cisplatin is one of the major chemotherapy medicines, and as such, its encapsulation or that of cisplatin derivatives has been frequently investigated in cages. Clever, Shionoya *et al.* have

successfully incorporated Magnus Salt derivatives (alternating stacks of positively and negatively charged square-planar Pt complexes)[72] into $Pt^{II}_2L_4$ cage **1.20** (Figure 1.14 b).[73]

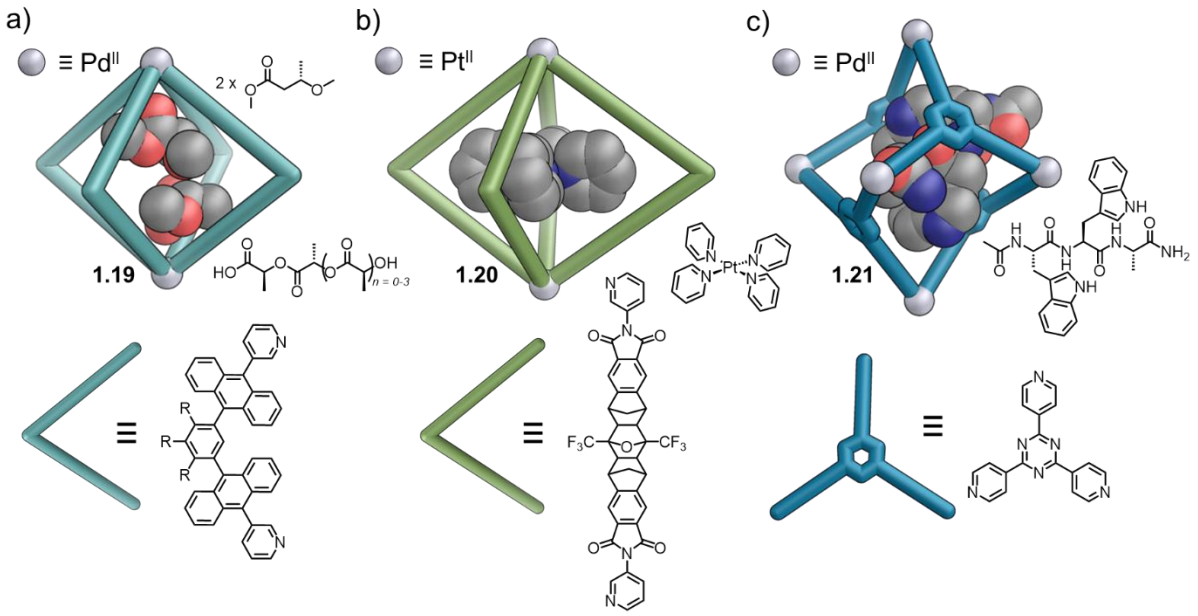


Figure 1.14 | Different coordination cages encapsulate biologically relevant molecules such as lactic acid derivatives (a),[70] platinum complexes (b)[73] or peptides (c).[71]

Furthermore, as a result of their generally hydrophobic cavity, cages have the ability to increase the solubility of molecules that would otherwise be insoluble in aqueous media. As an example, the co-encapsulation of otherwise water-insoluble Ir and Rh-Cp-type metal complexes along with aromatic compounds has been achieved in water thanks to the stabilisation effect of an M_6L_4 metallacage.[74] Therrien *et al.* have shown the encapsulation and solubilisation of pyrenylcycloplatinate complexes in water soluble $Ru_6L_2L'_3$ cages **1.22**, demonstrating the increased solubility of guests.[75] They have also focused on the incorporation of either $(acac)_2Pt$ or $(acac)_2Pd$ complexes in **1.22**, which can be toxic to cancerous cells (Figure 1.15). The cytotoxicities of the drug-in-cage complexes were compared to the drugs by themselves, showing acute toxicity of the complexes which were described to acts as a "Trojan horse", facilitating the uptake of the drug in cells.[76]

The positive charges induced by metal ions or positive groups on coordination cages can also balance out negative charges common in drug molecules, leading to higher levels of encapsulation. This strategy has been successfully implemented to improve the penetration of compounds in cells. In two cases, Mascarenas *et al.* used a fully organic positively charged cage to facilitate the internalisation of both pyranine containing peptides[77] and gold-

nanoparticles in cells.[78] Combining the benefits of both increased overall charge and increased solubility, Lippard *et al*. used a Pt_6L_4 cage to bind four platinum prodrugs that could then penetrate cells and later be activated by ascorbic acid.[79]

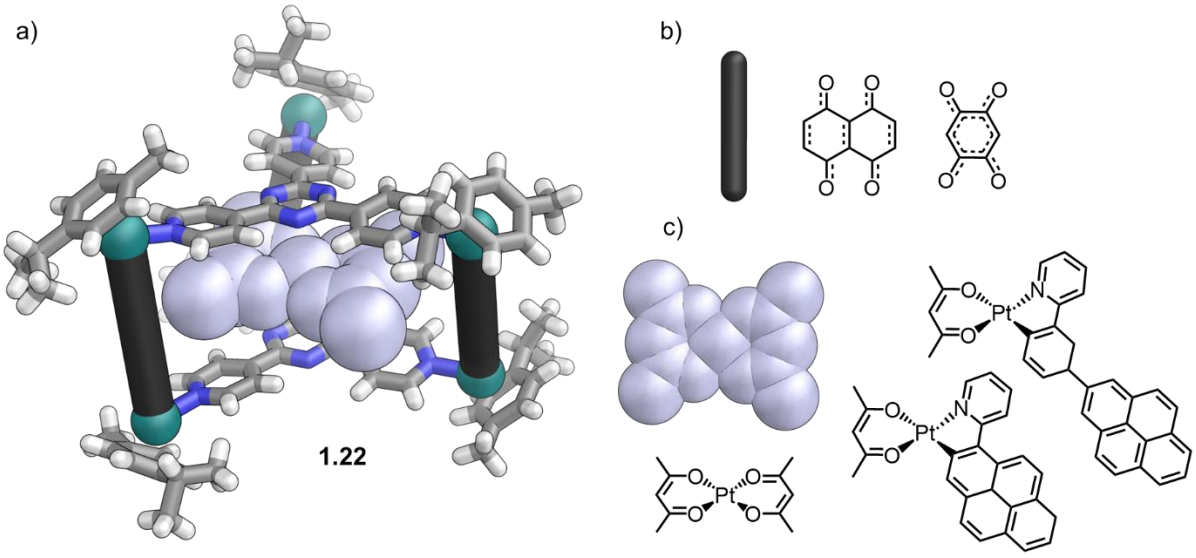


Figure 1.15 | a) Crystal structure of $(acac)_2Pt \subset Ru_6L_2L'_3$ cage **1.22**, where the black stick can be either of ligands dioxydo-1,4-naphthoquinonato or 2,5-dihydroxy-1,4-benzoquinonato (b). Cage **1.22** binds the various Pt complexes shown in c.[76]

Another important factor for drug delivery is the potential to release the encapsulated drug-molecule. To do so, triggers are commonly used to disassemble the cage or render the host-guest complex unfavourable. Light has largely been explored as a way to release guests by incorporating a photoswitching moiety in the cage ligands' backbone. Upon light irradiation, the change of conformation imposed on the cage leads to contraction of the cavity and the expulsion of the guest.[80]

Changing the pH of the bulk solution has also been a popular approach for guest release. Severin *et al*. used a photoacid generator to induce protonation of a range of pyridyl-containing ligands, leading to disassembly of the cages.[81] On the other hand, Hunter, Ward *et al*. exploited protonation of the guest molecules to induce release.[82] A collection of guest molecules bearing acidic or basic groups were encapsulated in Co_8L_{12} cage **1.23** which was stable over a range of pH. Depending on the pKa of the guest, protonation of the guest occurred at different pH values, leading to their decomplexation from the cage's cavity (Figure 1.16).

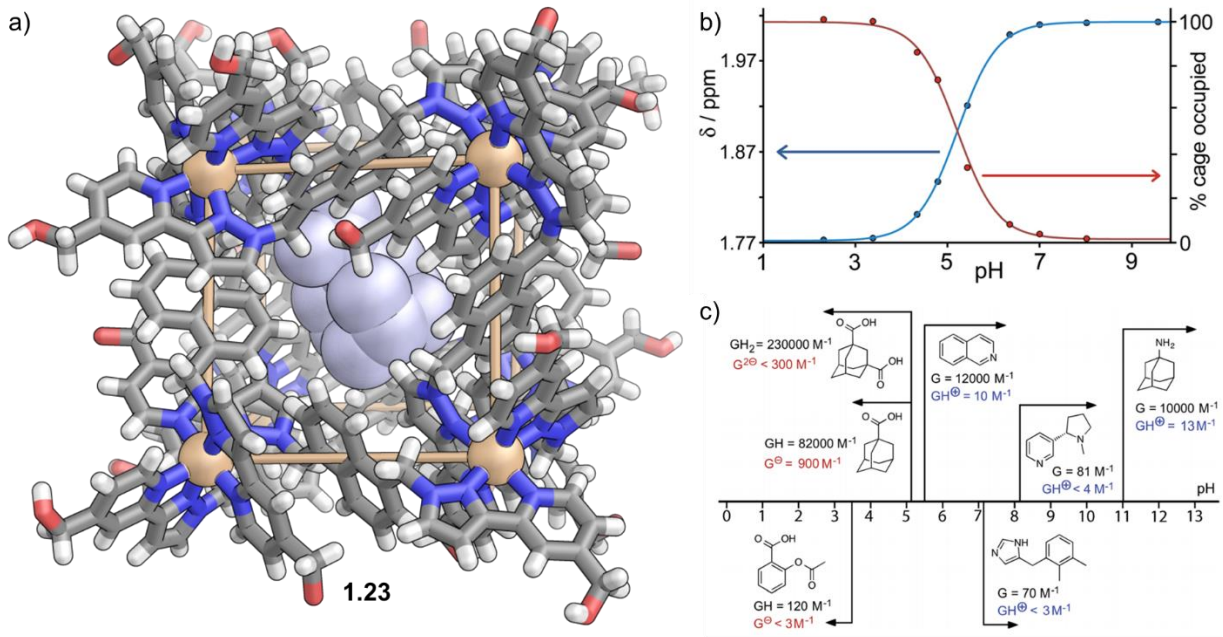


Figure 1.16 | a) Crystal structure of 1-adamantane-carboxylic acid ⊂ Co_8L_{12} **1.23**. b) Chemical shift of one of the signals of 1-adamantane-carboxylic acid (blue curve) and occupancy of the cavity of **1.23** (red curve) as a function of pH. c) Association constants for guests in neutral and charged states.[82] Figures adapted with permission from reference 82.

1.3 Biomolecules and self-assembly

Initial work by E. Fisher in the late 19th century described the "lock and key" principle between an enzyme and a substrate,[83] setting the platform for the study of weak interactions, molecular recognition, and host-guest chemistry from which supramolecular chemistry emerged. Natural molecules like peptides, proteins, nucleotides, carbohydrates and the interactions that allow these to cooperate and form higher ordered structures have been inspirational to chemists. Since then, chemists have tried to imitate the confined space found in enzymes to perform chemical reactions or reproduce host-guests interactions. Combining biomolecules and man-made supramolecular architectures represent the next logical step to better mimic and understand nature's marvels.

1.3.1 Supramolecular cages and biological systems

Supramolecular architectures are bridging the gap between small molecules and larger complex biomolecules (or assemblies of biomolecules) such as proteins or membranes. As such, the study of the interactions between supramolecular coordination cages and biological systems

and the potential of relatively small cages to impact larger biomolecular assemblies has expanded recently.

One of the main focuses has been the use of supramolecular metal complexes in biology as imaging agents[84-85] and therapeutics.[79] To this end, Hannon *et al.* have extensively investigated the interaction of M_2L_3 helicates with DNA,[86] RNA[87] and proteins.[88] They demonstrated that coordination cages can bind specifically within the grove of certain DNA and RNA architectures like three-way junctions (Figure 1.17). Building on these results, the versatile helicate **1.24** was shown to act as a potential anti-cancer drug[89] or inhibitor for viral replication in HIV by binding to the TAT protein.[90]

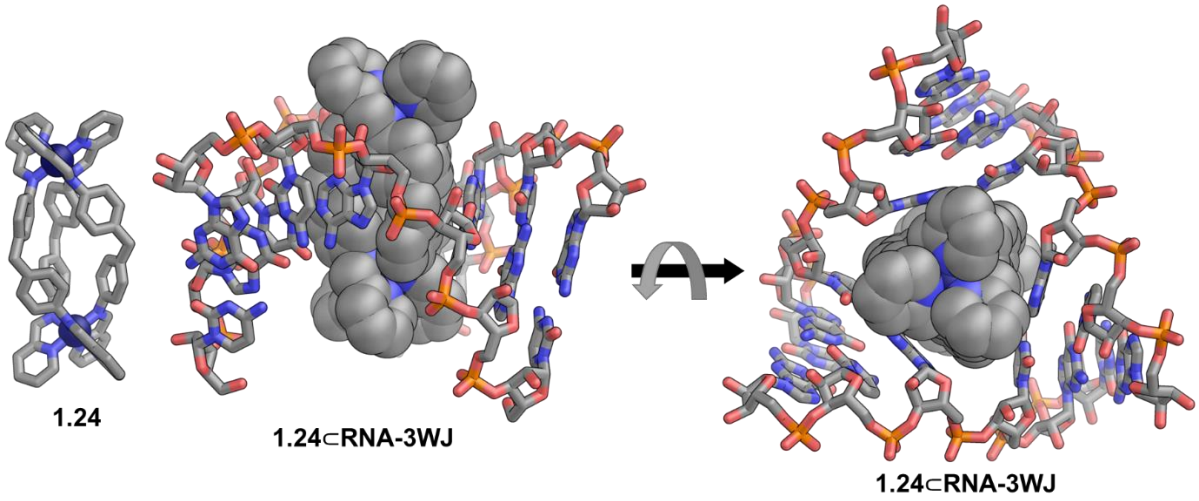


Figure 1.17 Crystal structure of $Fe^{II}_2L_3$ **1.24**, and **1.24** bound within a RNA three-way junction (view from the side and the top).[87]

Others have investigated the binding of coordination cages to transmembrane nanopores[91] or even the potential to use them as transmembrane pores.[92] The binding of single molecules to a tetrahedral coordination cage could be detected by monitoring the ion current through a α-hemolysin transmembrane protein which the cage interacted with, extending the scope for nanopore-detection strategies.[93]

1.3.2 Biomolecular cages

Incorporating biomolecules as building blocks into supramolecular cages has proven a successful strategy to render these biocompatible. In this context, the field of DNA nanocages has expanded recently, thanks to the increased control gained in DNA synthesis. The nanocages formed have good programmability and responsiveness and thus represent advantageous

carriers in applications like cargo delivery or biosensing and bioimaging.[94] Forming cages out of proteins and peptides has also been a popular approach.[95] The different polar and non-polar environments found on a peptide backbone provides great host-guest properties and recognition. Furthermore, the peptide cages' ability to mimic their biological compartments is now providing a way towards protocells for example.[96]

Although cages made of proteins,[97] peptides[98] or DNA[99] are common, small synthetic coordination cages incorporating biomolecules in their framework are rare. One of the few examples is the formation of a *β*-barrel demonstrated by Fujita *et al.* through metal directed self-assembly leading to folding of the peptide-based ligand.[100] An octapeptide combining both a *β*-strand and a loop-forming sequence was designed and self-assembled with either ZnI_2, $ZnCl_2$ or $ZnBr_2$, forming a pore bearing barrel made of three Zn_2L_2 macrocyles (Figure 1.18). The design employed here mimics the tertiary and quaternary structures only found in nature so far.

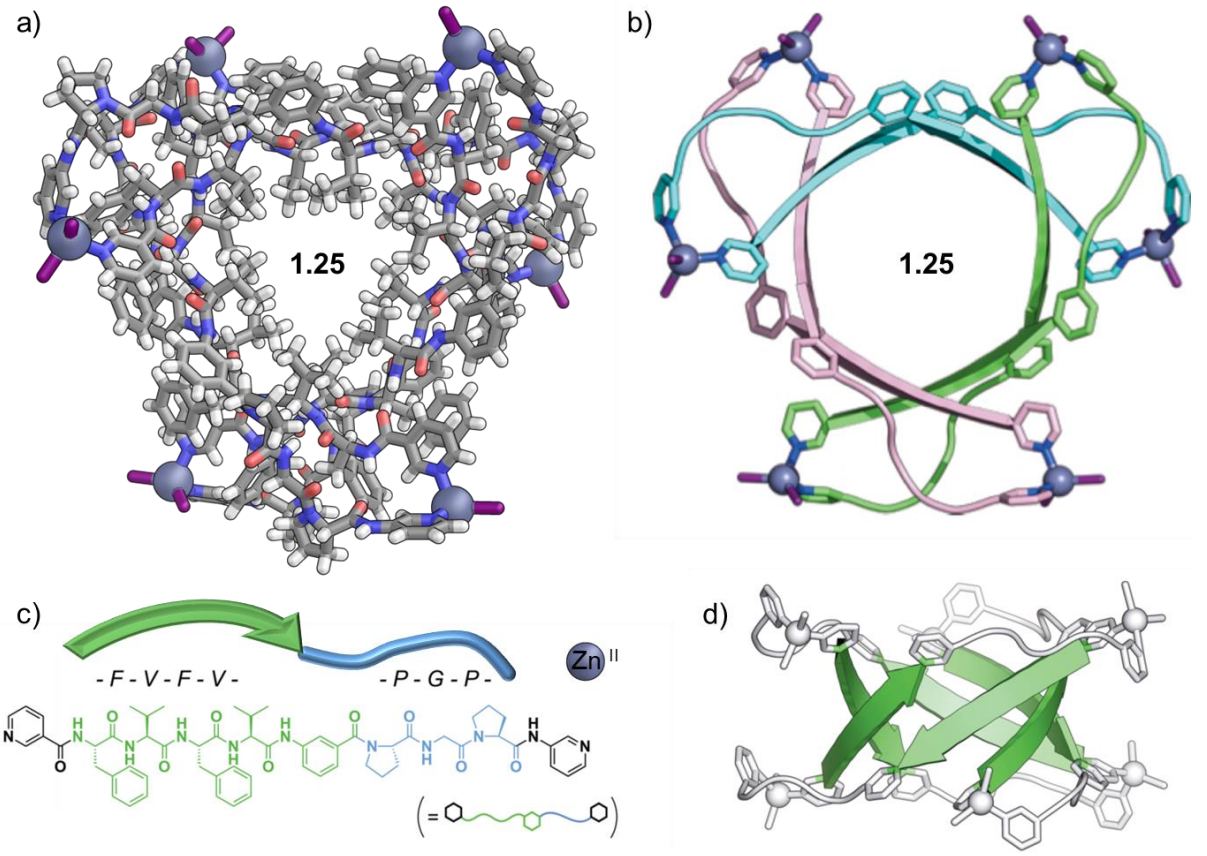


Figure 1.18 | a) Crystal structure (top view) of *β*-barrel **1.25** made from three macrocyclic $Zn^{II}_2L_2$ units and b) the cartoon representation with each $Zn^{II}_2L_2$ unit in a different colour. c) Structure of the components used: octapeptide highlighting the barrel forming sequence (green arrow, F-V-F-V) and the loop forming sequence (blue string, P-G-P) and Zn^{II} (purple sphere). d) Side view of schematic representation of **1.25** with *β*-strands highlighted in green.[100] Figures adapted with permission from reference 100.

Coordination has also been used in supramolecular structures as a way to stabilize the architecture formed from biomolecules. The incorporation of glycol based pyridine ligands into oligonucleotide strands was investigated by Clever *et al*. as a way to stabilize DNA quadruplexes for example. More precisely, when a pyridyl ligand was integrated into telomeric DNA sequences (repeating TTAGGG units), a G-quadruplex was formed with one of the tetrad replaced by the ligand. Upon binding of Cu, the folding of the quadruplex was enhanced by thermal stabilization.[101]

1.3.3 Bioconjugated self-assembled cages

Since the formation of synthetic coordination structures from biomolecules remains quite challenging, a simpler approach relies on tagging biomolecules at the surface of cages, while maintaining the original, better understood framework of the supramolecular cage. This strategy can potentially transform the coordination cage into a biocompatible vehicle without modifying the central cavity and therefore the host-guest properties of the cage.

Pioneering work in this field was carried out by Fujita *et al.* as early as the mid-2000s when he functionalised the inner cavity of a $Pd_{12}L_{24}$ capsule with peptides through amide coupling, leading to the generation of a chiral environment within the cavity.[102] A similar strategy was used to graft the outside of a self-assembled coordination cage with mono-, di-, tri- or poly-saccharides (Figure 1.19 a) which could undergo further higher order assembly by aggregation, mediated by either proteins[103] or Ca^{+} ions.[104] Bioconjugation of DNA fragments was also achieved, showing the overall structure could bind complementary oligonucleotides at its periphery.[105]

More recently, Casini *et al*. investigated two strategies for attaching peptides to M_2L_4 supramolecular cages. Bioconjugation of the peptide pre-assembly to the cage was shown to be a more successful strategy than post-assembly modification of the cage.[106] This method was used to perform the exo-biofunctionalisation of $Pd^{II}{}_2L_4$ **1.27** with four different integrin binding ligands (Figure 1.19 b).[107] After encapsulation of cisplatin in the cage's cavity, both increased anticancer activity and decreased toxicity on healthy cells were observed.

However, only a few examples describing biotagged supramolecular coordination containers are reported to date. Despite the advantages of coordination cages, their application to the

biomedical field has been far less explored than other systems such as MOFs,[108] nanoparticles[109] or liposomes.[110]

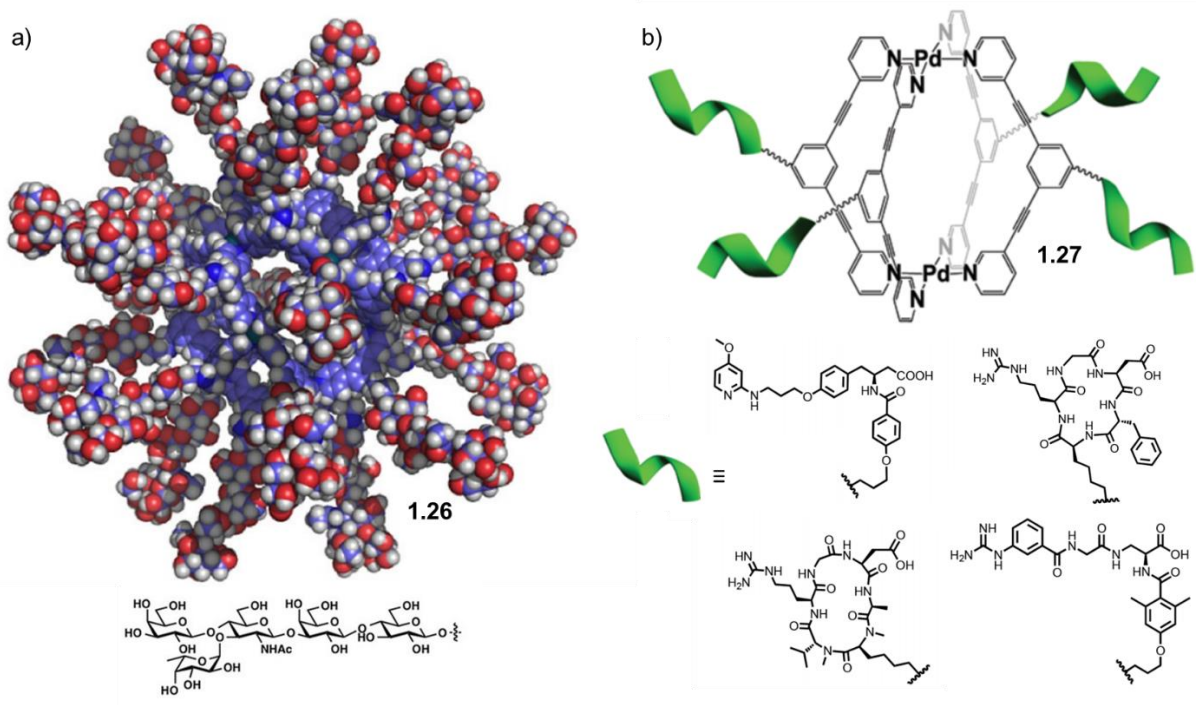


Figure 1.19 | Examples of bioconjugated coordination cages **1.26** and **1.27** functionalised respectively with a) saccharides, here a Lewis X type sugar,[104] or b) peptides, here integrin ligands.[107] Figures adapted with permission from references 104 and 107.

1.4 Supramolecular self-assembled gels

Supramolecular interactions have played a fundamental role in the development of molecule-based soft-materials. Amongst those, self-assembled gels have gained an increasing popularity over the years, with applications in fields as diverse as biomaterials, environmental remediation or self-healing materials to cite only a few.[111]

1.4.1 Low Molecular Weight Gels

Unlike the more common gels formed by polymeric compounds, most supramolecular gels are made of smaller molecules, known as low molecular weight gelators (LMWG), which are held together by non-covalent linkages. A high level of control can be gained thanks to the ease of design of the small molecules used to form the gels, leading to new properties on the macroscopic scale.

The gelation process of LMWG relies on the self-association of molecules to form long, polymer-like fibrous aggregates, which get entangled forming a matrix which traps the solvent, forming the bulk material (Figure 1.20).[112] As a consequence, supramolecular gels must be studied at all scales, using a wide range of methods such as NMR or UV-vis spectroscopy on the molecular level, microscopy (TEM, SEM, confocal) on the fibril level (nanostructures) or rheology on the macroscopic level.

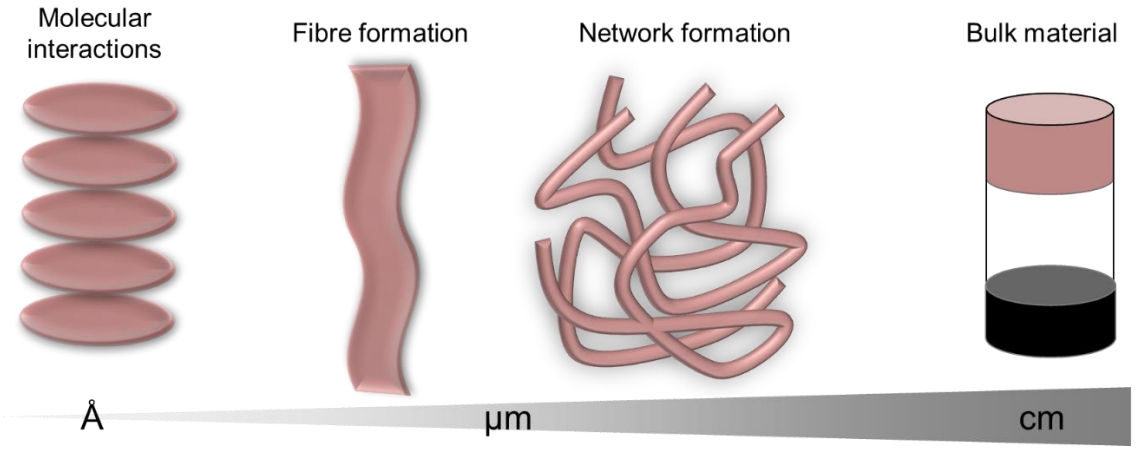


Figure 1.20 | Supramolecular gels form from assembly occurring at different scales, from molecular interactions to fibres, to network and finally bulk material.

The desired physical properties can be enforced on the bulk material by fine-tuning of the molecular gelator properties. For example, changes of the gel aspect have been observed upon application of an external stimulus such as UV light. Jung, Lee, Kim *et al*. have shown the possibility to perform reversible lithography (write and erase) on a gel made of tris(4-((E)-phenyldiazenyl)phenyl)-benzene-1,3,5-tricarboxamide by localised melting of the fibres.[113] As a popular trigger, UV light was also employed by Smith *et al*. to promote the formation of patterned gel domains made of DBS-gly **1.29** (DBS=1,3:2,4-dibenzylidene-D-sorbitol, opaque gel) within a preformed gel matrix made of DBS-CO_2H **1.28** (translucent gel).[114] Due to the differences in pH of these molecules (pKa = 5.4 for **1.28** and pKa = 4.3 for **1.29**), the selective gelation of **1.29** could be obtained by pH change initiated by UV irradiation of diphenyliodonium nitrate (DPIN, Figure 1.21).

Furthermore, the reversible nature of the interactions holding the gel together most often render the crosslinking and the gelation reversible as well. Self-healing can therefore be easily achieved by LMWG, as demonstrated by the multitude of examples in the literature.[115] Stimuli responsiveness is aided by the reversible nature of the bonds in the gel. A sol-gel transition[116]

in LMWG was shown to be obtained *via* redox,[117] light,[118] heat,[119] sound[120] or combination of these stimuli.[121]

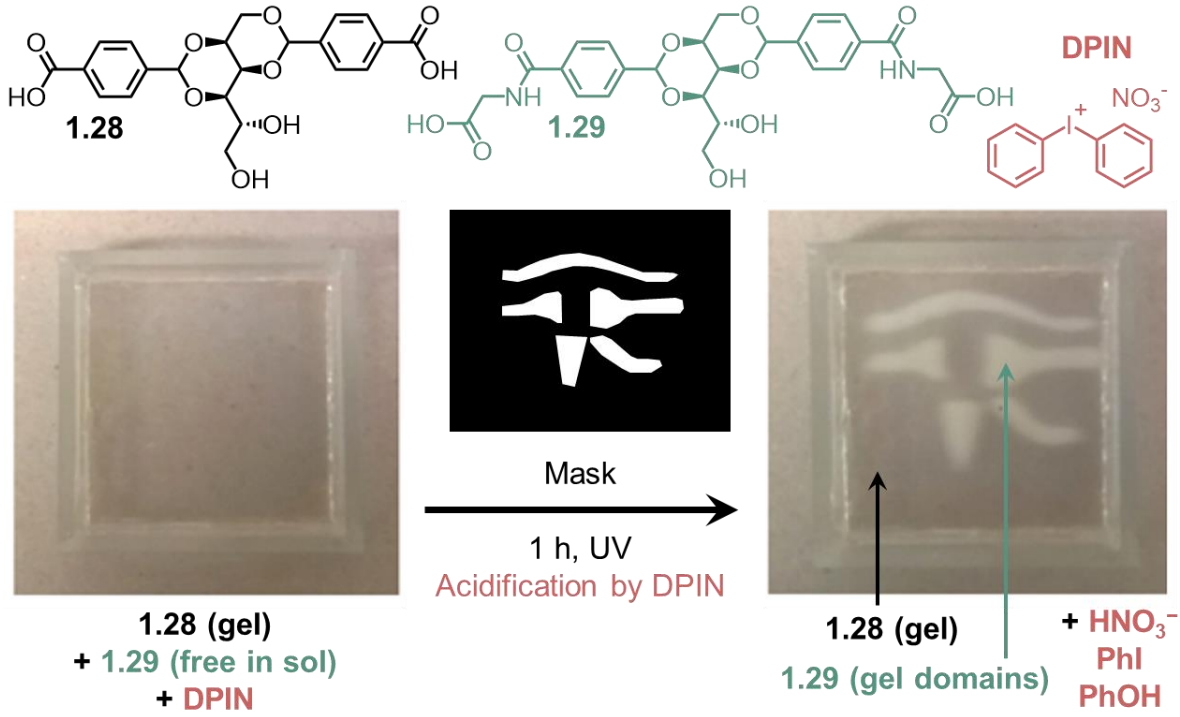


Figure 1.21 | Photo-patterning of a multi-domain gel containing DBS-CO_2H (**1.28**) and a DBS-gly (**1.29**). First a gel forms made of fibrils containing exclusively **1.28**. Following acidification by UV irradiation through a mask, a pattern is imprinted *via* formation of gel domains rich in **1.29**.[114] Figures adapted with permission from reference 114.

1.4.2 Self-assembled gels from short peptides

Naturally occurring small molecules, peptides have versatile biological properties and have been used widely to form peptide-based materials. Their propensity to aggregate and form stacks or fibrils makes small peptides ideal candidates for the formation of supramolecular gels. Easily obtained by solid-phase synthesis, a wide variety of peptides ranging from only a few amino acids (our focus here) to longer sequences up to dozens of amino acids have been studied for their gel-forming ability and biocompatibility.[122-124]

Very small peptides were discovered to aggregate into defined nano-structures, the simplest to date being the tube-forming diphenylalanine (Phe-Phe).[125] Other dipeptides have been shown to form gels, but only when the N-terminus was protected by an Fmoc (fluorenylmethyloxycarbonyl) moiety.[126] Fmoc-dipeptides and mainly Fmoc-Phe-Phe are

amongst the most extensively studied LMWG due to their excellent mechanical properties. However, identifying natural, non-substituted peptides able to form gels is of great interest.

With this objective in mind, Ulijn, Tuttle *et al.* carried out a computational study screening 8000 tri-peptide combinations (Figure 1.22).[127] A set of design rules that promote aggregation were drawn from the systematic study, showing that aromatic amino acids (with apolar groups in blue in the MARTINI code, Figure 1.22 a) are most favourable in the middle or end of a tripeptide, while positive and hydrogen-bonding (yellow or orange) residues should be placed at the N-terminus and negative (red) residues in position at the C-terminus.

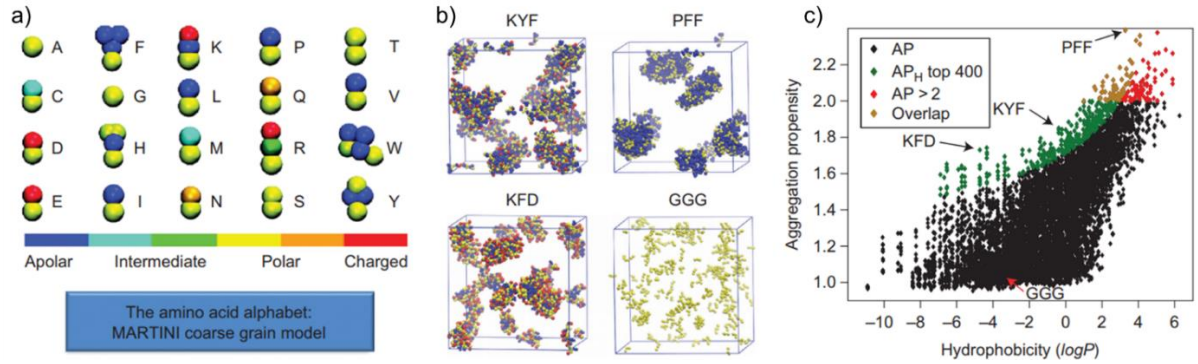


Figure 1.22 | Screening for self-assembling tripeptides. a) Representation of amino acids in the MARTINI force field. b) Simulation showing the aggregation for four tripeptides: KYF, KFD, PFF and GGG. c) Aggregation propensity (AP) as a function of hydrophobicity for 8,000 tripeptides. Tripeptides with AP > 2 are denoted in red, the top 400 tripeptides from the APH score (APH = hydrophobicity-corrected AP, taking into account the solubility of the peptide as well) are given in green with the overlapping candidates shown in orange. Arrows point to data points for KYF, KFD, PFF and GGG.[127] Figures adapted with permission from reference 126.

Formation of nano-structures designed to favour the stacking of tripeptides was achieved by incorporating amino acids of alternating chirality. The first example dates back to 1993 when nanotubes were obtained from octapeptides of alternating L and D residues.[128] Subsequently, Marchesan *et al.* have applied this strategy to the formation of gels from heterochiral tripeptides.[129]

Hydrophobic peptides of general structure Phe-DXxx-Phe, with Xxx being valine or leucine isomers, were shown to assemble into fibrils, unlike the less hydrophobic sequences tested.[130] The nanostructures observed were diverse in terms of rigidity, branching or tendency to form bundles, as shown by TEM and cryo-TEM (Figure 1.23). The assembly into structures was explained by bending of the heterochiral tripeptide backbones to maximize non-covalent interactions and exclude water from specific regions, resulting in zipping of the amino-acids and aggregation.

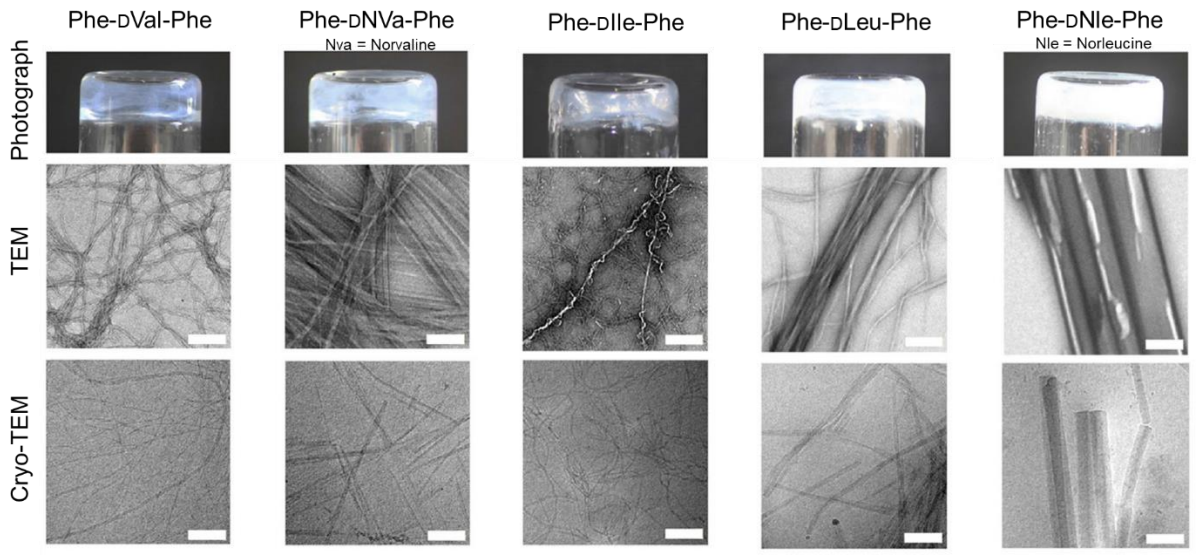


Figure 1.23 | Photograph, TEM and cryo-TEM images of hydrogels obtained from alternating chirality tripeptide Phe-DXxx-Phe, with DXxx = valine, norvaline, isoleucine, leucine or norleucine.[130] Figures adapted with permission from reference 129.

1.4.3 Coordination cages in gels

Metal coordination has been widely used to promote the crosslinking of polymeric strands, forming the matrix of gels.[131] The same principle applies to the incorporation of coordination cages in gels with the advantage that the higher nuclearity of the cage allows for increased branching of the polymer network.

The most common approach to obtaining gels containing cages is to append cage-forming building blocks onto soluble polymer strands. Upon coordination with metal ions, the cages assemble, acting as the nodes of the gel matrix. A typical example of this strategy was reported by Johnson *et al*. Bent di-pyridyl ligands were connected by PEG linkers, which upon addition of $Pd(NO_3)_2$ and heating at 80 °C formed the polymer/cage system.[132] Based on the ligand used (Figure 1.24 f), either small M_2L_4 cages, or larger $M_{12}L_{24}$ cages were formed (Figure 1.24 a), leading to various degrees of crosslinking and therefore changes in the mechanical properties of the material. However, in the case where the large cage was formed, a high number of loop-defects, where the linker connected two points on the same cage, were observed (shown in red in Figure 1.24). This effect was limited by lengthening the linker, thus decreasing the amount of loop obtained (Figure 1.24 e).[133]

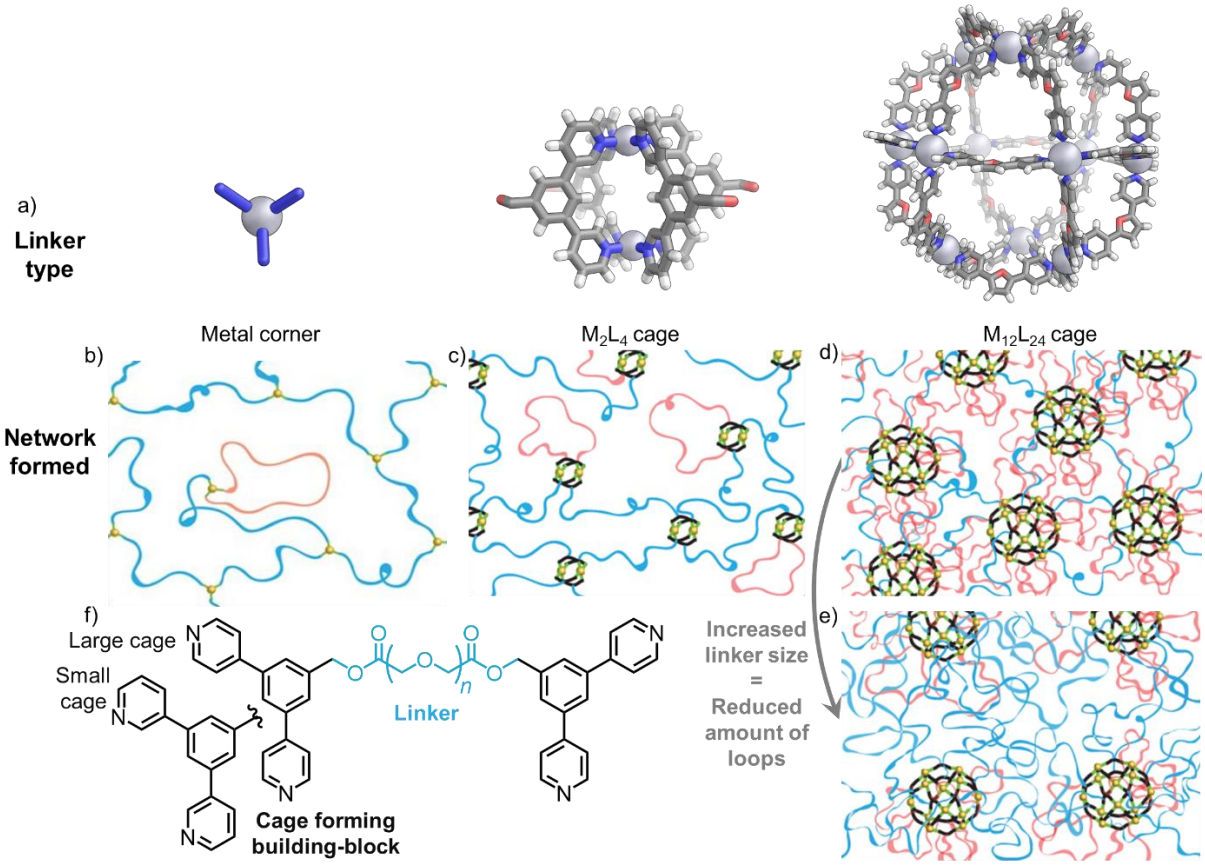


Figure 1.24 | a) Representation of the complexes used as linkers, here a metal corner, a M_2L_4 or a $M_{12}L_{24}$ cage. Schematic representation of the polymeric networks obtained upon formation of metal complexes (b), small M_2L_4 cages (c), or larger $M_{12}L_{24}$ cages with either short (d) or long (e) linkers. f) Polymeric linker used with variable length PEG chain and variable cage-forming building blocks.[132-133] Figures adapted with permission from reference 131 and 132.

High levels of control over the gel properties directly results from the shape and size of the cage chosen, leading to well defined polymers with regular cavities and intrinsic microporosity.[134] The incorporation of coordination cages into gels also allows the combination of both the properties of the cage and gel, yielding extremely versatile materials.[135] For example, Stang *et al.* demonstrated the formation of a fluorescent supramolecular gel. The fluorescence was induced by a tetraphenylethene bearing ligand while the gel network was formed *via* host-guest interactions between the crown ether grafted on the cage's framework and bis-ammonium linkers.[136]

The very sought after self-healing property is facilitated in self-assembled coordination gels due to the reversible nature of the chemical bonds present. Fujita *et al.* described a $M_{12}L_{14}$ cage, grafted with mesogenic arms which self-assembled into spherical liquid crystals. When combined with 1) the same cage-core grafted with aldehydes and 2) a diamine crosslinker, formation of a hybrid liquid crystal-gel with self-healing properties was triggered.[137] Stimuli-

responsive self-healing was demonstrated by Johnson *et al.*[138] Light-responsive cages which switch between M_3L_6 and $M_{24}L_{48}$[139] were crosslinked by PEG chains. Due to the difference in kinetics of the ligand exchange, the gel formed by the M_3L_6 linkage proved to be self-healable, while that made of the $M_{24}L_{48}$ cage was not.

However, the geometry of the cage is not the only important factor. The properties of gels can also be controlled based on the shape of the linker. Maji *et al.* nicely demonstrated the importance of this parameter by gaining control over the gel nano-structures through the linker binding between cubic M_8L_{12} cages **1.30** (Figure 1.25 a).[140] When combined with either N-(2-aminoethyl)-1,3-propanediamine, guanidine.HCl or β-alanine, cage **1.30** assembled to form nanobouquets, nanosheets or nanocubes respectively, thus tuning the properties of the hydrogel on the nano-, micro- and macroscopic scale (Figure 1.25 d-g).

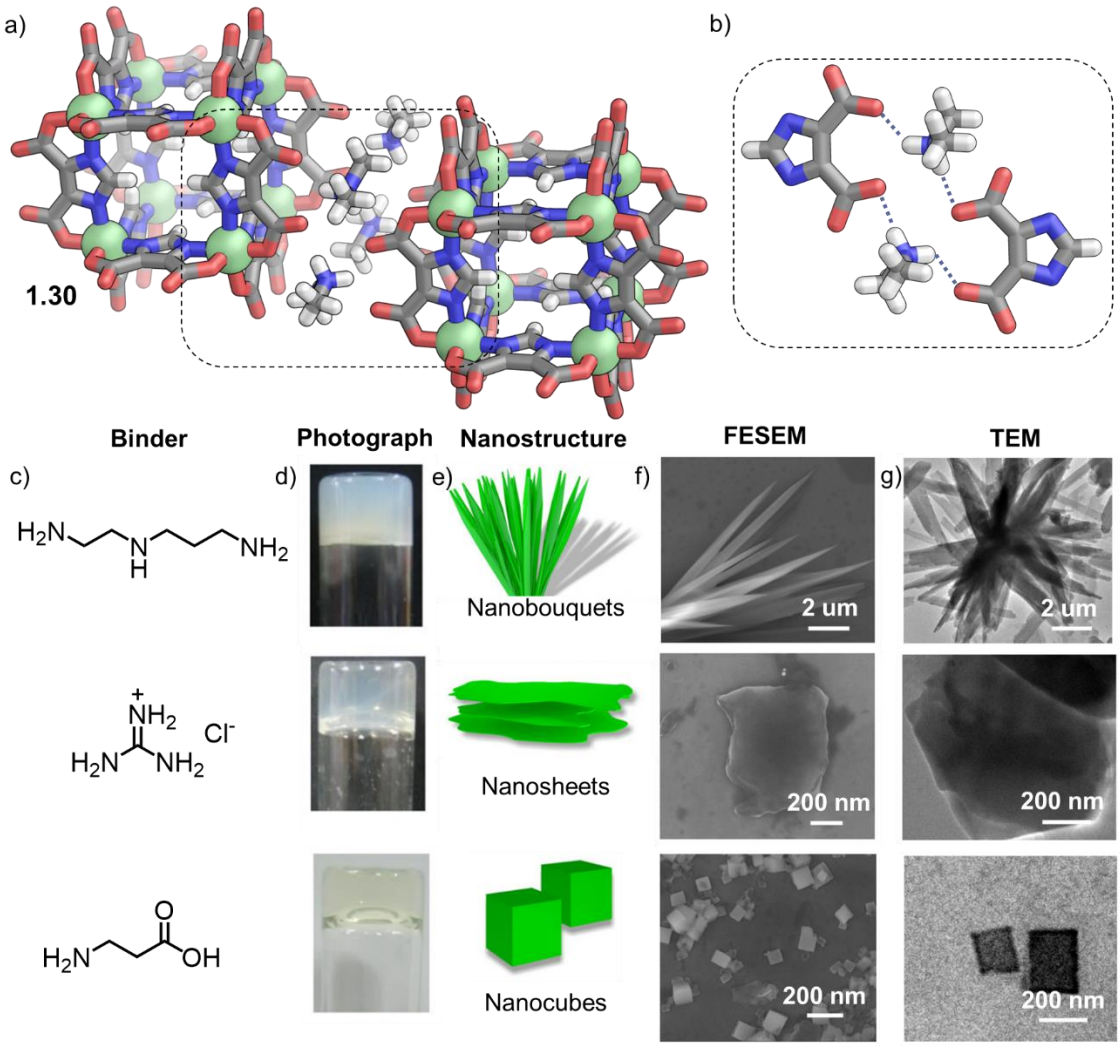


Figure 1.25 | a) Crystal structure of cage **1.30** interacting with multiple $Me_2NH_2^+$ cations. b) Portion of the crystal structure showing ligands from two different cages connected via intermolecular H-bonding with four $Me_2NH_2^+$ cations. Gels (d) formed when **1.30** was combined with different binders (c) and the FESEM and TEM images, revealing the gel nanostructures (e).[140]

1.5 Aims and overview

The combination of coordination cages with biomolecules could lead to new properties arising either from each component – cage and biomolecules – separately, or emerging from the synergy of the two systems. This thesis will present a series of investigations into the interactions of sub-component self-assembled cages and biomolecules. More precisely, cages will be tested as a way to influence the structure of peptide LMWG and as a way to increase the gel selectivity towards guests. Functionalisation of coordination cages by direct covalent attachment of biomolecules, tri-peptides and sugars, will also be probed, showing new properties arising from this unique combination.

With the rise of more complex, biologically relevant chemical systems, the study of multicomponent supramolecular self-sorting is a necessity. Understanding the driving forces which lead molecules to sort into precise well defined structures will enable the design of composite synthetic systems closer in performance to those of nature. However, the investigation of complex systems has been challenging and limited by current analytical methods. With this goal in mind, part of this thesis will be dedicated to the investigation of mass-spectrometry as a quantitative tool for the study of self-sorted systems. Three different chemical systems were assessed, showing that quantitative information on the structural preferences of the systems could be gained from simple MS analysis.

1.6 References

[1] R. A. Bilbeisi, J.-C. Olsen, L. J. Charbonnière, A. Trabolsi, *Inorg. Chim. Acta* **2014**, *417*, 79-108.

[2] T. R. Cook, P. J. Stang, *Chem. Rev.* **2015**, *115*, 7001-7045.

[3] D. L. Caulder, K. N. Raymond, *J. Chem. Soc., Dalton Trans.* **1999**, 1185-1200.

[4] M. Yoshizawa, J. K. Klosterman, *Chem. Soc. Rev.* **2014**, *43*, 1885-1898.

[5] T. R. Cook, Y.-R. Zheng, P. J. Stang, *Chem. Rev.* **2013**, *113*, 734-777.

[6] R. Chakrabarty, P. S. Mukherjee, P. J. Stang, *Chem. Rev.* **2011**, *111*, 6810-6918.

[7] H. Yokoyama, Y. Ueda, D. Fujita, S. Sato, M. Fujita, *Chem. Asian J.* **2015**, *10*, 2292-2295.

[8] K. Suzuki, M. Tominaga, M. Kawano, M. Fujita, *Chem. Commun.* **2009**, 1638-1640.

[9] M. Tominaga, K. Suzuki, M. Kawano, T. Kusukawa, T. Ozeki, S. Sakamoto, K. Yamaguchi, M. Fujita, *Angew. Chem. Int. Ed.* **2004**, *43*, 5621-5625.

[10] Q.-F. Sun, J. Iwasa, D. Ogawa, Y. Ishido, S. Sato, T. Ozeki, Y. Sei, K. Yamaguchi, M. Fujita, *Science* **2010**, *328*, 1144-1147.

[11] D. Fujita, Y. Ueda, S. Sato, H. Yokoyama, N. Mizuno, T. Kumasaka, M. Fujita, *Chem* **2016**, *1*, 91-101.

[12] D. H. Busch, *Acc. Chem. Res.* **1978**, *11*, 392-400.

[13] J. Hamblin, L. J. Childs, N. W. Alcock, M. J. Hannon, *J. Chem. Soc., Dalton Trans.* **2002**, 164-169.

[14] D. Schultz, J. R. Nitschke, *Proc. Nat. Acad. Sci. U. S. A.* **2005**, *102*, 11191-11195.

[15] J. R. Nitschke, *Angew. Chem. Int. Ed.* **2004**, *43*, 3073-3075.

[16] J. R. Nitschke, *Acc. Chem. Res.* **2007**, *40*, 103-112.

[17] T. K. Ronson, S. Zarra, S. P. Black, J. R. Nitschke, *Chem. Commun.* **2013**, *49*, 2476-2490.

[18] D. A. Roberts, A. M. Castilla, T. K. Ronson, J. R. Nitschke, *J. Am. Chem. Soc.* **2014**, *136*, 8201-8204.

[19] A. Jiménez, R. A. Bilbeisi, T. K. Ronson, S. Zarra, C. Woodhead, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2014**, *53*, 4556-4560.

[20] F. J. Rizzuto, W.-Y. Wu, T. K. Ronson, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2016**, *55*, 7958-7962.

[21] W. J. Ramsay, T. K. Ronson, J. K. Clegg, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2013**, *52*, 13439-13443.

[22] S. Zarra, J. K. Clegg, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2013**, *52*, 4837-4840.

[23] J. L. Greenfield, F. J. Rizzuto, I. Goldberga, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2017**, *56*, 7541-7545.

[24] S. P. Black, A. R. Stefankiewicz, M. M. J. Smulders, D. Sattler, C. A. Schalley, J. R. Nitschke, J. K. M. Sanders, *Angew. Chem. Int. Ed.* **2013**, *52*, 5749-5752.

[25] R. A. Bilbeisi, T. K. Ronson, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2013**, *52*, 9027-9030.

[26] M. E. Belowich, J. F. Stoddart, *Chem. Soc. Rev.* **2012**, *41*, 2003-2024.

[27] Y. R. Hristova, M. M. J. Smulders, J. K. Clegg, B. Breiner, J. R. Nitschke, *Chem. Sci.* **2011**, *2*, 638-641.

[28] M. Kieffer, B. S. Pilgrim, T. K. Ronson, D. A. Roberts, M. Aleksanyan, J. R. Nitschke, *J. Am. Chem. Soc.* **2016**, *138*, 6813-6821.

[29] D. A. Roberts, B. S. Pilgrim, J. D. Cooper, T. K. Ronson, S. Zarra, J. R. Nitschke, *J. Am. Chem. Soc.* **2015**, *137*, 10068-10071.

[30] E. G. Percástegui, J. Mosquera, T. K. Ronson, A. J. Plajer, M. Kieffer, J. R. Nitschke, *Chem. Sci.* **2019**, *10*, 2006-2018.

[31] E. G. Percástegui, J. Mosquera, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2017**, *56*, 9136-9140.

[32] A. B. Grommet, J. B. Hoffman, E. G. Percástegui, J. Mosquera, D. J. Howe, J. L. Bolliger, J. R. Nitschke, *J. Am. Chem. Soc.* **2018**, *140*, 14770-14776.

[33] S. L. Dabb, N. C. Fletcher, *Dalton Trans.* **2015**, *44*, 4406-4422.

[34] I. A. Riddell, M. M. J. Smulders, J. K. Clegg, Y. R. Hristova, B. Breiner, J. D. Thoburn, J. R. Nitschke, *Nat. Chem.* **2012**, *4*, 751-756.

[35] J. L. Bolliger, A. M. Belenguer, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2013**, *52*, 7958-7962.

[36] T. Tateishi, T. Kojima, S. Hiraoka, *Commun. Chem.* **2018**, *1*, 1-12.

[37] T. Liu, Y. Liu, W. Xuan, Y. Cui, *Angew. Chem.* **2010**, *122*, 4215-4218.

[38] W. Meng, J. K. Clegg, J. D. Thoburn, J. R. Nitschke, *J. Am. Chem. Soc.* **2011**, *133*, 13652-13660.

[39] M. M. Safont-Sempere, G. Fernández, F. Würthner, *Chem. Rev.* **2011**, *111*, 5784-5814.

[40] Z. He, W. Jiang, C. A. Schalley, *Chem. Soc. Rev.* **2015**, *44*, 779-789.

[41] W. M. Bloch, G. H. Clever, *Chem. Commun.* **2017**, *53*, 8506-8516.

[42] Q.-F. Sun, S. Sato, M. Fujita, *Angew. Chem. Int. Ed.* **2014**, *53*, 13510-13513.

[43] T. K. Ronson, D. A. Roberts, S. P. Black, J. R. Nitschke, *J. Am. Chem. Soc.* **2015**, *137*, 14502-14512.

[44] P. Howlader, P. Das, E. Zangrando, P. S. Mukherjee, *J. Am. Chem. Soc.* **2016**, *138*, 1668-1676.

[45] W. M. Bloch, Y. Abe, J. J. Holstein, C. M. Wandtke, B. Dittrich, G. H. Clever, *J. Am. Chem. Soc.* **2016**, *138*, 13750-13755.

[46] T. Sawada, M. Yoshizawa, S. Sato, M. Fujita, *Nat. Chem.* **2009**, *1*, 53.

[47] F. Rizzuto, J. Carpenter, J. Nitschke, *Unpublished results*.

[48] Y. B. Voloshin, Irina; Krämer, Roland, in *The Encapsulation Phenomenon: Synthesis, Reactivity and Applications of Caged Ions and Molecules.*, Springer International Publishing., **2016**.

[49] S. Mecozzi, J. Rebek, *Chem. Eur. J.* **1998**, *4*, 1016-1022.

[50] S. Turega, W. Cullen, M. Whitehead, C. A. Hunter, M. D. Ward, *J. Am. Chem. Soc.* **2014**, *136*, 8475-8483.

[51] T. K. Ronson, C. Giri, N. Kodiah Beyeh, A. Minkkinen, F. Topić, J. J. Holstein, K. Rissanen, J. R. Nitschke, *Chem.—Eur. J.* **2013**, *19*, 3374-3382.

[52] A. S. Singh, S.-S. Sun, *RSC Adv.* **2012**, *2*, 9502-9510.

[53] W. Meng, B. Breiner, K. Rissanen, J. D. Thoburn, J. K. Clegg, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2011**, *50*, 3479-3483.

[54] S. Mirtschin, A. Slabon-Turski, R. Scopelliti, A. H. Velders, K. Severin, *J. Am. Chem. Soc.* **2010**, *132*, 14004-14005.

[55] T. K. Ronson, W. Meng, J. R. Nitschke, *J. Am. Chem. Soc.* **2017**, *139*, 9698-9707.

[56] P. Ballester, *Chem. Soc. Rev.* **2010**, *39*, 3810-3830.

[57] Julia L. Brumaghim, M. Michels, D. Pagliero, Kenneth N. Raymond, *Eur. J. Org. Chem.* **2004**, *2004*, 5115-5118.

[58] D. A. Roberts, B. S. Pilgrim, G. Sirvinskaite, T. K. Ronson, J. R. Nitschke, *J. Am. Chem. Soc.* **2018**, *140*, 9616-9623.

[59] D. Zhang, T. K. Ronson, J. Mosquera, A. Martinez, L. Guy, J. R. Nitschke, *J. Am. Chem. Soc.* **2017**, *139*, 6574-6577.

[60] I. A. Riddell, Y. R. Hristova, J. K. Clegg, C. S. Wood, B. Breiner, J. R. Nitschke, *J. Am. Chem. Soc.* **2013**, *135*, 2723-2733.

[61] R. L. Paul, Z. R. Bell, J. C. Jeffery, J. A. McCleverty, M. D. Ward, *Proc. Nat. Acad. Sci.* **2002**, *99*, 4883-4888.

[62] R. Custelcean, P. V. Bonnesen, N. C. Duncan, X. Zhang, L. A. Watson, G. Van Berkel, W. B. Parson, B. P. Hay, *J. Am. Chem. Soc.* **2012**, *134*, 8525-8534.

[63] N. Busschaert, C. Caltagirone, W. Van Rossom, P. A. Gale, *Chem. Rev.* **2015**, *115*, 8038-8155.

[64] Z. Liu, M. Frasconi, J. Lei, Z. J. Brown, Z. Zhu, D. Cao, J. Iehl, G. Liu, A. C. Fahrenbach, Y. Y. Botros, O. K. Farha, J. T. Hupp, C. A. Mirkin, J. Fraser Stoddart, *Nat. Commun.* **2013**, *4*, 1855.

[65] J.-R. Li, J. Sculley, H.-C. Zhou, *Chem. Rev.* **2012**, *112*, 869-932.

[66] D. Zhang, T. K. Ronson, J. Mosquera, A. Martinez, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2018**, *57*, 3717-3721.

[67] X.-Z. Li, L.-P. Zhou, L.-L. Yan, Y.-M. Dong, Z.-L. Bai, X.-Q. Sun, J. Diwu, S. Wang, J.-C. Bünzli, Q.-F. Sun, *Nat. Commun.* **2018**, *9*, 547.

[68] V. P. Torchilin, in *Drug Deliv.* (Ed.: M. Schäfer-Korting), Springer Berlin Heidelberg, Berlin, Heidelberg, **2010**, pp. 3-53.

[69] T. Lammers, F. Kiessling, W. E. Hennink, G. Storm, *J. Controlled Release* **2012**, *161*, 175-187.

[70] S. Kusaba, M. Yamashina, M. Akita, T. Kikuchi, M. Yoshizawa, *Angew. Chem. Int. Ed.* **2018**, *57*, 3706-3710.

[71] S. Tashiro, M. Tominaga, M. Kawano, B. Therrien, T. Ozeki, M. Fujita, *J. Am. Chem. Soc.* **2005**, *127*, 4546-4547.

[72] M. Atoji, J. W. Richardson, R. E. Rundle, *J. Am. Chem. Soc.* **1957**, *79*, 3017-3020.

[73] G. H. Clever, W. Kawamura, S. Tashiro, M. Shiro, M. Shionoya, *Angew. Chem. Int. Ed.* **2012**, *51*, 2606-2609.

[74] S. H. A. M. Leenders, R. Becker, T. Kumpulainen, B. de Bruin, T. Sawada, T. Kato, M. Fujita, J. N. H. Reek, *Chem. Eur. J.* **2016**, *22*, 15468-15474.

[75] N. P. E. Barry, O. Zava, W. Wu, J. Zhao, B. Therrien, *Inorg. Chem. Commun.* **2012**, *18*, 25-28.

[76] B. Therrien, G. Süss-Fink, P. Govindaswamy, A. K. Renfrew, P. J. Dyson, *Angew. Chem. Int. Ed.* **2008**, *47*, 3773-3776.

[77] J. Rodríguez, J. Mosquera, J. R. Couceiro, J. R. Nitschke, M. E. Vázquez, J. L. Mascareñas, *J. Am. Chem. Soc.* **2017**, *139*, 55-58.

[78] J. Mosquera, M. Henriksen-Lacey, I. García, M. Martínez-Calvo, J. Rodríguez, J. L. Mascareñas, L. M. Liz-Marzán, *J. Am. Chem. Soc.* **2018**, *140*, 4469-4472.

[79] Y.-R. Zheng, K. Suntharalingam, T. C. Johnstone, S. J. Lippard, *Chem. Sci.* **2015**, *6*, 1189-1193.

[80] M. Han, R. Michel, B. He, Y.-S. Chen, D. Stalke, M. John, G. H. Clever, *Angew. Chem. Int. Ed.* **2013**, *52*, 1319-1323.

[81] S. M. Jansze, G. Cecot, K. Severin, *Chem. Sci.* **2018**, *9*, 4253-4257.

[82] W. Cullen, S. Turega, C. A. Hunter, M. D. Ward, *Chem. Sci.* **2015**, *6*, 625-631.

[83] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985-2993.

[84] L. Fang, G. Trigiante, C. J. Kousseff, R. Crespo-Otero, M. P. Philpott, M. Watkinson, *Chem. Commun.* **2018**, *54*, 9619-9622.

[85] B. P. Burke, W. Grantham, M. J. Burke, G. S. Nichol, D. Roberts, I. Renard, R. Hargreaves, C. Cawthorne, S. J. Archibald, P. J. Lusby, *J. Am. Chem. Soc.* **2018**, *140*, 16877-16881.

[86] D. R. Boer, J. M. C. A. Kerckhoffs, Y. Parajo, M. Pascu, I. Usón, P. Lincoln, M. J. Hannon, M. Coll, *Angew. Chem. Int. Ed.* **2010**, *49*, 2336-2339.

[87] S. Phongtongpasuk, S. Paulus, J. Schnabl, R. K. O. Sigel, B. Spingler, M. J. Hannon, E. Freisinger, *Angew. Chem. Int. Ed.* **2013**, *52*, 11513-11516.

[88] L. Cardo, I. Nawroth, P. J. Cail, J. A. McKeating, M. J. Hannon, *Scientific Reports* **2018**, *8*, 13342.

[89] G. I. Pascu, A. C. G. Hotze, C. Sanchez-Cano, B. M. Kariuki, M. J. Hannon, *Angew. Chem. Int. Ed.* **2007**, *46*, 4374-4378.

[90] J. Malina, M. J. Hannon, V. Brabec, *Scientific Reports* **2016**, *6*, 29674.

[91] S. Matile, A. Vargas Jentzsch, J. Montenegro, A. Fin, *Chem. Soc. Rev.* **2011**, *40*, 2453-2474.

[92] C. J. E. Haynes, P. A. Gale, *Chem. Commun.* **2011**, *47*, 8203-8209.

[93] S. Borsley, J. A. Cooper, P. J. Lusby, S. L. Cockroft, *Chem. Eur. J.* **2018**, *24*, 4542-4546.

[94] K. E. Bujold, A. Lacroix, H. F. Sleiman, *Chem* **2018**, *4*, 495-521.

[95] D. Mandal, A. Nasrolahi Shirazi, K. Parang, *Org. Biomol. Chem* **2014**, *12*, 3544-3561.

[96] S. Bhaskar, S. Lim, *Npg Asia Materials* **2017**, *9*, e371.

[97] J. G. Heddle, S. Chakraborti, K. Iwasaki, *Current Opinion in Structural Biology* **2017**, *43*, 148-155.

[98] E. Bartolami, J. Knoops, Y. Bessin, M. Fossépré, J. Chamieh, P. Dumy, M. Surin, S. Ulrich, *Chem. Eur. J.* **2017**, *23*, 14323-14331.

[99] N. C. Seeman, H. F. Sleiman, *Nature Reviews Materials* **2017**, *3*, 17068.

[100] M. Yamagami, T. Sawada, M. Fujita, *J. Am. Chem. Soc.* **2018**, *140*, 8644-8647.

[101] D. M. Engelhard, J. Nowack, G. H. Clever, *Angew. Chem. Int. Ed.* **2017**, *56*, 11640-11644.

[102] K. Suzuki, M. Kawano, S. Sato, M. Fujita, *J. Am. Chem. Soc.* **2007**, *129*, 10652-10653.

[103] N. Kamiya, M. Tominaga, S. Sato, M. Fujita, *J. Am. Chem. Soc.* **2007**, *129*, 3816-3817.

[104] G. Yan, T. Yamaguchi, T. Suzuki, S. Yanaka, S. Sato, M. Fujita, K. Kato, *Chemistry – An Asian Journal* **2017**, *12*, 968-972.

[105] T. Kikuchi, S. Sato, M. Fujita, *J. Am. Chem. Soc.* **2010**, *132*, 15930-15932.

[106] J. Han, A. Schmidt, T. Zhang, H. Permentier, G. M. M. Groothuis, R. Bischoff, F. E. Kühn, P. Horvatovich, A. Casini, *Chem. Commun.* **2017**, *53*, 1405-1408.

[107] J. Han, A. F. B. Räder, F. Reichart, B. Aikman, M. N. Wenzel, B. Woods, M. Weinmüller, B. S. Ludwig, S. Stürup, G. M. M. Groothuis, H. P. Permentier, R. Bischoff, H. Kessler, P. Horvatovich, A. Casini, *Bioconjugate Chem.* **2018**, *29*, 3856-3865.

[108] M.-X. Wu, Y.-W. Yang, *Adv. Mater.* **2017**, *29*, 1606134.

[109] A. Z. Wilczewska, K. Niemirowicz, K. H. Markiewicz, H. Car, *Pharmacological Reports* **2012**, *64*, 1020-1037.

[110] L. Sercombe, T. Veerati, F. Moheimani, S. Y. Wu, A. K. Sood, S. Hua, *Frontiers in Pharmacology* **2015**, *6*.

[111] D. B. Amabilino, D. K. Smith, J. W. Steed, *Chem. Soc. Rev.* **2017**, *46*, 2404-2420.

[112] N. M. Sangeetha, U. Maitra, *Chem. Soc. Rev.* **2005**, *34*, 821-836.

[113] S. Lee, S. Oh, J. Lee, Y. Malpani, Y.-S. Jung, B. Kang, J. Y. Lee, K. Ozasa, T. Isoshima, S. Y. Lee, M. Hara, D. Hashizume, J.-M. Kim, *Langmuir* **2013**, *29*, 5869-5877.

[114] D. J. Cornwell, O. J. Daubney, D. K. Smith, *J. Am. Chem. Soc.* **2015**, *137*, 15486-15492.

[115] S. Strandman, X. X. Zhu, *Gels* **2016**, *2*, 16.

[116] C. D. Jones, J. W. Steed, *Chem. Soc. Rev.* **2016**, *45*, 6546-6596.

[117] C. Wang, D. Zhang, D. Zhu, *J. Am. Chem. Soc.* **2005**, *127*, 16372-16373.

[118] J. H. Kim, M. Seo, Y. J. Kim, S. Y. Kim, *Langmuir* **2009**, *25*, 1761-1766.

[119] G. Du, Y. Peng, Y. Pei, L. Zhao, Z. Wen, Z. Hu, *Energy & Fuels* **2017**, *31*, 9283-9289.

[120] X. Cai, Y. Wu, L. Wang, N. Yan, J. Liu, X. Fang, Y. Fang, *Soft Matter* **2013**, *9*, 5807-5814.

[121] S. Wang, W. Shen, Y. Feng, H. Tian, *Chem. Commun.* **2006**, 1497-1499.

[122] N. Habibi, N. Kamaly, A. Memic, H. Shafiee, *Nano Today* **2016**, *11*, 41-60.

[123] J. P. Jung, J. Z. Gasiorowski, J. H. Collier, *Peptide Science* **2010**, *94*, 49-59.

[124] K. Sato, M. P. Hendricks, L. C. Palmer, S. I. Stupp, *Chem. Soc. Rev.* **2018**, *47*, 7539-7551.

[125] M. Reches, E. Gazit, *Science* **2003**, *300*, 625-627.

[126] Y. Zhang, H. Gu, Z. Yang, B. Xu, *J. Am. Chem. Soc.* **2003**, *125*, 13680-13681.

[127] P. W. J. M. Frederix, G. G. Scott, Y. M. Abul-Haija, D. Kalafatovic, C. G. Pappas, N. Javid, N. T. Hunt, R. V. Ulijn, T. Tuttle, *Nat. Chem.* **2014**, *7*, 30-37.

[128] M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee, N. Khazanovich, *Nature* **1993**, *366*, 324-327.

[129] S. Marchesan, L. Waddington, C. D. Easton, D. A. Winkler, L. Goodall, J. Forsythe, P. G. Hartley, *Nanoscale* **2012**, *4*, 6752-6760.

[130] A. M. Garcia, D. Iglesias, E. Parisi, K. E. Styan, L. J. Waddington, C. Deganutti, R. De Zorzi, M. Grassi, M. Melchionna, A. V. Vargiu, S. Marchesan, *Chem* **2018**, *4*, 1862-1876.

[131] H. Li, P. Yang, P. Pageni, C. Tang, *Macromol. Rapid Commun.* **2017**, *38*, 1700109.

[132] A. V. Zhukhovitskiy, M. Zhong, E. G. Keeler, V. K. Michaelis, J. E. P. Sun, M. J. A. Hore, D. J. Pochan, R. G. Griffin, A. P. Willard, J. A. Johnson, *Nat. Chem.* **2015**, *8*, 33.

[133] A. V. Zhukhovitskiy, J. Zhao, M. Zhong, E. G. Keeler, E. A. Alt, P. Teichen, R. G. Griffin, M. J. A. Hore, A. P. Willard, J. A. Johnson, *Macromolecules* **2016**, *49*, 6896-6902.

[134] A. Carné-Sánchez, G. A. Craig, P. Larpent, T. Hirose, M. Higuchi, S. Kitagawa, K. Matsuda, K. Urayama, S. Furukawa, *Nat. Commun.* **2018**, *9*, 2506.

[135] P. Sutar, T. K. Maji, *Chem. Commun.* **2016**, *52*, 8055-8074.

[136] C. Lu, M. Zhang, D. Tang, X. Yan, Z. Zhang, Z. Zhou, B. Song, H. Wang, X. Li, S. Yin, H. Sepehrpour, P. J. Stang, *J. Am. Chem. Soc.* **2018**, *140*, 7674-7680.

[137] J. Uchida, M. Yoshio, S. Sato, H. Yokoyama, M. Fujita, T. Kato, *Angew. Chem. Int. Ed.* **2017**, *56*, 14085-14089.

[138] Y. Gu, E. A. Alt, H. Wang, X. Li, A. P. Willard, J. A. Johnson, *Nature* **2018**, *560*, 65-69.

[139] M. Han, Y. Luo, B. Damaschke, L. Gómez, X. Ribas, A. Jose, P. Peretzki, M. Seibt, G. H. Clever, *Angew. Chem. Int. Ed.* **2016**, *55*, 445-449.

[140] P. Sutar, V. M. Suresh, K. Jayaramulu, A. Hazra, T. K. Maji, *Nat. Commun.* **2018**, *9*, 3587.

Chapter 2

Materials and method

2.1 General

All reagents and solvents, unless stated otherwise, were purchased from commercial suppliers and were used without purification. Commercially supplied pararosaniline base **6.I** was found to contain impurities (including fuchsin). The material supplied by Aldrich was found to contain the least amount of this contaminant and could be obtained in satisfactory purity after multiple washes with CH_3CN. Trialdehyde **3.A**,[1] trianiline **3.B**,[2] dialdehyde **6.C**,[3] perfluorinated dialdehyde **6.D**,[4] iron(II)bis(trifluoromethane)sulfonimide,[5] and iron(II)hexafluorophosphate,[5] were prepared following literature procedures. Cobalt(II)bis(trifluoromethane)sulfonimide[6] and cobalt(II)(trifluoromethane)sulfonate,[6] 2-formylphenantroline **6.G**,[7] and trianline **6.J**[8] were synthesised by Felix Rizzuto following literature procedures. Trianiline **5.N** was synthesised by Dawei Zhang following a literature procedure.[9]

2.2 Mass spectrometry (MS)

Low resolution electrospray ionisation mass spectrometry (LR-ESI-MS) was carried out on a Micromass Quattro LC mass spectrometer (cone voltage 15-25 eV, desolvation temp. 313 K, ionization temp. 313 K) infused from a Harvard Syringe Pump at a rate of 10 μL min^{-1}. High resolution electrospray ionisation mass spectrometry (HR-ESI-MS) was undetaken on a Thermofisher LTQ Orbitrap XL hybrid ion trap mass spectrometer. Travelling Wave Ion Mobility Quadrupole Time-of-Flight mass spectra (IM-MS) were collected on a Waters Vion IMS QTof mass spectrometer equipped with XS Ion Optics and the QuanTof2 detection system.

2.3 Nuclear Magnetic Resonance (NMR) spectroscopy

NMR spectra were recorded at 298 K on a Bruker Avance III HD 400 MHz Smart Probe Spectrometer, a Bruker 500 MHz DCH Cryoprobe Spectrometer, a Bruker Avance III HD 500MHz Smart Probe Spectrometer or a Bruker Avance 500 MHz TCI Cryoprobe Spectrometer. Chemical shift (δ) values are reported in ppm relative to the residual solvent peaks for ^1H and ^{13}C and relative to an internal standard of C_6F_6 in CD_3CN ($\delta = -164.90$ ppm) for ^{19}F. Coupling constants (*J*) are reported in hertz (Hz) and the signal multiplicities are

described as s (singlet), d (doublet), dd (doublet of doublets), t (triplet), dt (doublet of triplets), m (multiplet) and b (broad).

Wide sweep paramagnetic NMR spectra were recorded in the analogue digitisation mode with a spectral width (SW) of 407.42 ppm, a transmitter frequency offset (O1P) of 130.00 ppm and an acquisition time of 0.1 sec. Due to the experimental difficulties associated with collecting NMR data for ^{1}H nuclei with vastly different relaxation times, differences between measured and theoretical integration values were in some cases observed. While the paramagnetic nature of the complex precluded complete assignment of the proton environments, we propose that through-bond proximity of the proton environment to each Co^{II} centres dictates the extent of downfield shifting of each signal, as observed in previous reports.[6, 10]

DOSY NMR experiments were performed on a Bruker Avance 500 MHz TCI Cryoprobe Spectrometer. Maximum gradient strength was 5.3 G/cm A with a gradient highest current of 10 A. Gradient amplitudes ranged from 5 to 95%. The standard Bruker pulse program, ledbpgp2s, employing a stimulated echo and longitudinal eddy-current delay (LED) using bipolar gradient pulses for diffusion using 2 spoil gradients was utilised. The diffusion delay (d20) was set to 0.80 s and the diffusion gradient length (p30) to 500 µs.

2.4 UV-Vis spectroscopy

UV-Visible absorption spectroscopy was performed using an Agilent Cary 5000 UV-Vis-NIR spectrophotometer at 298 K. Measurements were background subtracted from blank solvent in an identical cuvette. Samples were analysed using quartz cuvettes with optical path lengths of 1, 10 or 50 mm.

2.5 Molecular modelling

Molecular model simulations (MM2 and MM3 force fields) of supramolecular complexes were performed using SCIGRESS version FJ 2.6 (EU 3.1.9) Build 5996.8255.20141202.

2.6 Fluorescence spectroscopy

Fluorescence measurements were performed on a Cary Eclipse Fluorescence spectrophotometer with a 10 mm path length quartz cuvette at 295 K.

2.7 Circular Dichroism (CD) spectroscopy

Circular Dichroism spectroscopy was performed on an Applied-Photophysics Chirascan qCD spectrometer using a 1 mm path-length quartz cuvette. Experiments were recorded at 298 K, maintained with a Peltier temperature control. Measurements were background subtracted from blank solvent in an identical cuvette. The sample concentrations were adjusted to maintain a HV below 700 Volts. A minimum sample integration time of 1 second was used. The data was smoothed using a 10 pts Savitzky-Golay algorithm and the residuals checked for distortions. CD spectra were acquired by Jake Greenfield.

2.8 Oscillatory rheometry

Dynamic time sweep rheological analyses were performed on a Malvern Kinexus Ultra Plus Rheometer with a 20 mm stainless steel parallel plate geometry. Experiments were recorded at 298 K, maintained with a Peltier temperature control. Each sample was prepared *in situ* and immediately analyzed with a gap of 1 mm. Time sweeps were recorded for 1 h, using a frequency of 2 Hz and a controlled stress of 2 Pa. After 1 h, frequency sweeps were recorded from 0.1 to 10 Hz using a controlled stress of 2 Pa. Finally, stress sweeps were recorded using a frequency of 2 Hz until the breaking point for every gel. Rheometry data was acquired by Ana García Fernández.

2.9 Transmission electron microscopy (TEM) and energy-dispersive X-ray (EDX) spectroscopy

TEM analyses were performed on JEM 2100 equipped with an Oxford Instruments INCS energy-dispersive X-ray spectroscopy (EDXS) detector operated at 100 kV. TEM grids (copper-grid-supported lacey carbon film) were first exposed to the UV-ozone cleaner (UV-Ozone Procleaner Plus) for 10 min. After deposition on the TEM grid, the samples were dried for 15 min at 298 K, and contrasted by aqueous tungsten phosphate solution (pH=7.4). TEM images and EDX spectra were acquired by Slavko Kralj.

2.10 References

[1] A. M. Castilla, N. Ousaka, R. A. Bilbeisi, E. Valeri, T. K. Ronson, J. R. Nitschke, *J. Am. Chem. Soc.* **2013**, *135*, 17999-18006.

[2] J. L. Bolliger, T. K. Ronson, M. Ogawa, J. R. Nitschke, *J. Am. Chem. Soc.* **2014**, *136*, 14545-14553.

[3] S. Ma, M. M. J. Smulders, Y. R. Hristova, J. K. Clegg, T. K. Ronson, S. Zarra, J. R. Nitschke, *J. Am. Chem. Soc.* **2013**, *135*, 5678-5684.

[4] M. Kieffer, B. S. Pilgrim, T. K. Ronson, D. A. Roberts, M. Aleksanyan, J. R. Nitschke, *J. Am. Chem. Soc.* **2016**, *138*, 6813-6821.

[5] M. P. Sibi, G. Petrovic, *Tetrahedron: Asymmetry* **2003**, *14*, 2879-2882.

[6] I. A. Riddell, M. M. J. Smulders, J. K. Clegg, Y. R. Hristova, B. Breiner, J. D. Thoburn, J. R. Nitschke, *Nat. Chem.* **2012**, *4*, 751-756.

[7] M. Zhang, R. Gao, X. Hao, W.-H. Sun, *J. Organomet. Chem.* **2008**, *693*, 3867-3877.

[8] A. Ferguson, M. A. Squire, D. Siretanu, D. Mitcov, C. Mathonière, R. Clérac, P. E. Kruger, *Chem. Commun.* **2013**, *49*, 1597-1599.

[9] D. Zhang, T. K. Ronson, J. Mosquera, A. Martinez, L. Guy, J. R. Nitschke, *J. Am. Chem. Soc.* **2017**, *139*, 6574-6577.

[10] C. S. Wood, T. K. Ronson, A. M. Belenguer, J. J. Holstein, J. R. Nitschke, *Nat. Chem.* **2015**, *7*, 354.

Chapter 3

Chemical separation performed by $Fe^{II}_4L_4$ tetrahedral cages incorporated in a tripeptide gel

3.1 Introduction

Systems relying on supramolecular interactions have generally been investigated in the solution state. However, some challenges such as chemical separation by supramolecular hosts would be better approached if bulk materials were to be used. Gels are materials that share properties of solids and liquids, and exhibit stimuli-responsiveness,[1] which enables applications that include environmental remediation[2] and cargo delivery.[3] The incorporation of metal complexes and hollow metal-organic cages into gels[4] can alter the responsiveness of the material to stimuli, thus endowing it with useful properties comprising fluorescence and self-healing. [5-6] Chemically-modified coordination cages have been reported to act as nodes within gels and polymers, with linkages between cages formed by polymeric chains or binders.[7-8] The ability of cages to selectively encapsulate molecular cargoes led to materials with high affinities for one compound or a family of compounds specifically.[9-10] However, synthetic challenges often arise from this approach due to the high degree of modifications necessary to obtain gels from molecular cages.

The identification of a suitable gel matrix, compatible with metalorganic cages could enable the immobilization of cages within the gel, without any chemical modification. Under the condition that the gel pores are small enough, the large 3D molecular cages could be entrapped in the matrix. Short peptides[11] are readily obtained, biocompatible materials[12] that form gels under a wide variety of conditions. More specifically, tripeptides of alternating L-D-L stereochemistry have recently been shown to adopt a conformation with side chains in an isotactic configuration that enables their long-range self-organization into gel-forming fibres (Figure 3.1).[13]

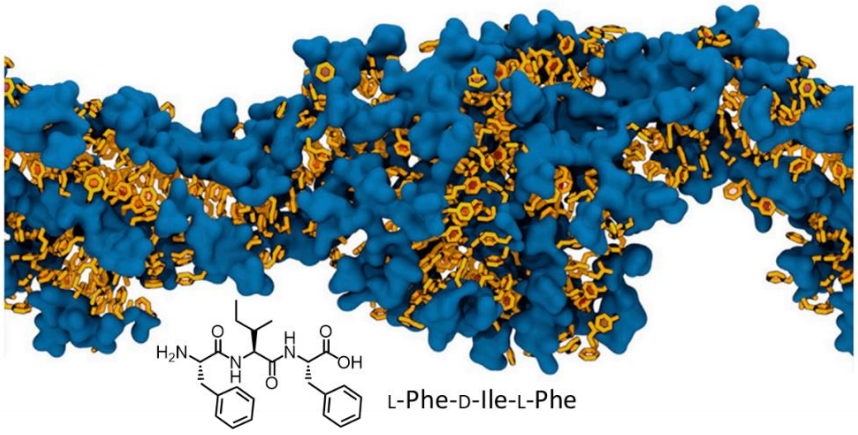


Figure 3.1 | L-Phe-D-Ile-L-Phe tripeptide generates fibres, wherein hydrophobic side chains (orange) are localised on the opposite side of the hydrophilic peptide backbone (blue).[14] Figure reproduced with permission from reference 14.

In this chapter, gels made from a new tripeptide LMWG (low molecular weight gelator) were prepared and used to separate and immobilize different cages within distinct regions, but within a single solvent phase. The peptide gel was assembled in the presence of existing self-assembled coordination cages, embedding these in its matrix formed from peptide fibrils. The hybrid gel material contained a homogenous array of cage cavities with functional guest-binding capabilities. The spatially-separated cages were shown to bind distinct small molecule guests, allowing these guests to be segregated within defined parts of the gel and thus separated from a mixture by selective encapsulation.

3.2 Immobilisation of cages in a tripeptide gel – Stability and properties of the material

The subcomponents **3.A** and **3.B** were synthesized according to literature procedures.[15-16] Self-assembly of $Fe(NTf_2)_2$ and either *p*-toluidine and **3.A** or picolinaldehyde and **3.B** yielded cages **3.1** and **3.2** respectively (Figure 3.2).[15, 17]

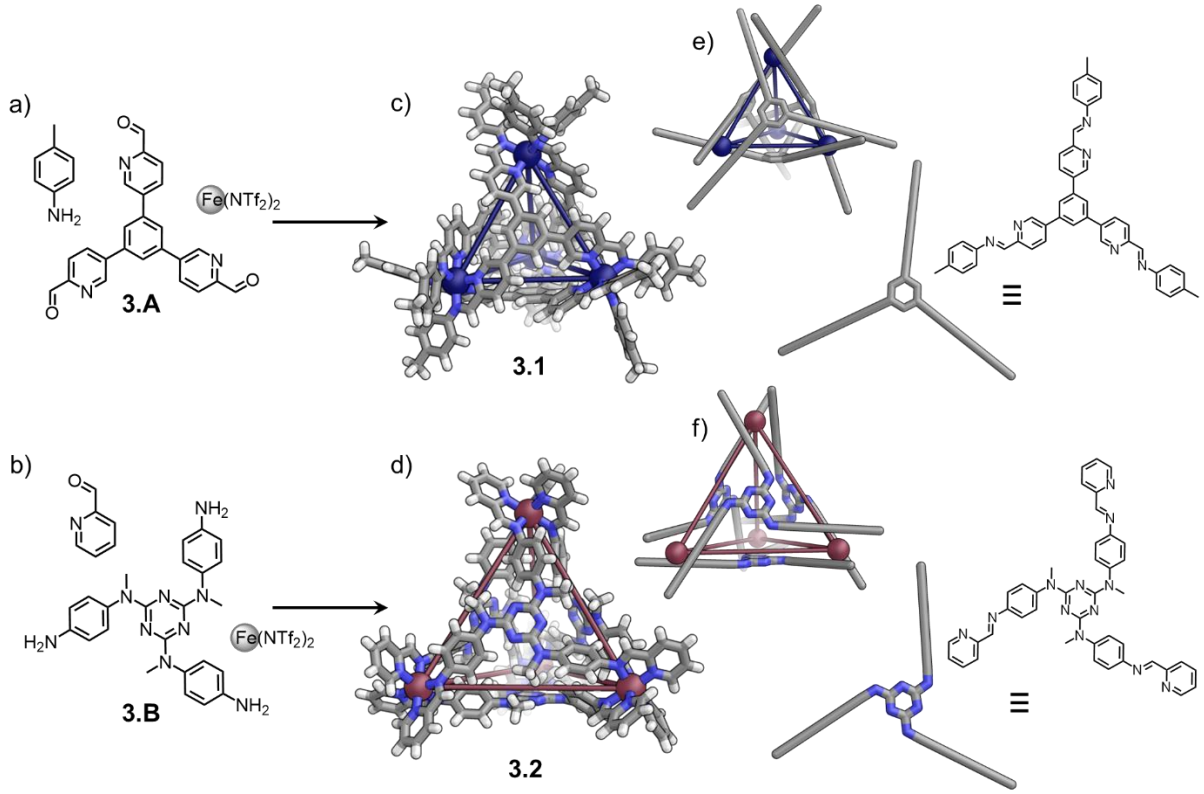


Figure 3.2 | Self-assembly of a) cage **3.1** from *p*-toluidine, **3.A** and $Fe(NTf_2)_2$ and b) cage **3.2** from picolinaldehyde, **3.B** and $Fe(NTf_2)_2$. c) and d) single crystal X-ray structures of **3.1** and **3.2** with the linkages between Fe^{II} centres (blue or pink spheres) added to highlight the frame. e) and f) schematic representations of **3.1** and **3.2** and their ligands.

Cage **3.1** encapsulated trifluoroacetate (TFA^-) in slow exchange, as evidenced by the changes observed in ^1H and ^{19}F NMR upon addition of 20 eq. of KTFA (notably a new signal corresponding to the encapsulated TFA^- appeared at –79.8 ppm in the ^{19}F NMR spectrum). Further addition of 1 eq. of $TBAReO_4$ resulted in the displacement of the encapsulated TFA^-, indicative of the stronger affinity between **3.1** and ReO_4^- compared to TFA^- (Figure 3.3 b and c). Cage **3.2** encapsulated NTf_2^- very weakly as evidence by the small second set of peaks observed in the ^1H NMR. As previously reported, **3.2** encapsulated fluoroadamantane (FA) sufficiently strongly to displace the encapsulated NTf_2^- (Figure 3.3 e and f).[18]

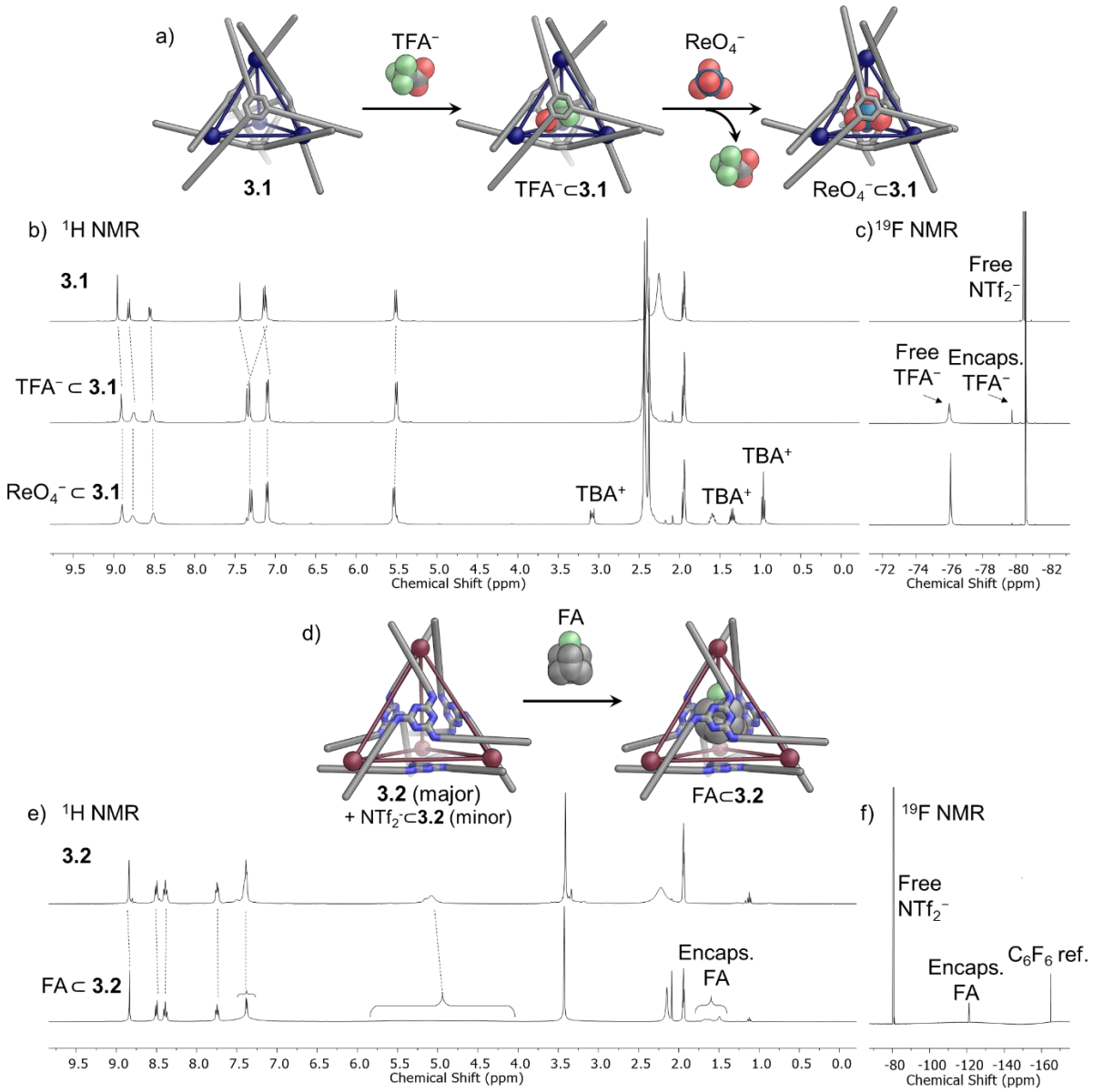


Figure 3.3 | Host-guest properties of a) **3.1** and d) **3.2**. b) and c) ^1H NMR spectra (400 MHz, 298 K, CD_3CN) and ^{19}F NMR spectra (376 MHz, 298K, CD_3CN) of **3.1** before and after the addition of 20 eq. of KTFA followed by 1 eq. of $TBAReO_4$. e) and f) ^1H NMR spectra (400 MHz, 298 K, CD_3CN) and ^{19}F NMR spectrum (376 MHz, 298K, CD_3CN) of **3.2** and FA⊂**3.2**.

Work in the Marchesan group is focused on the synthesis of small syndiotactic L-D-L tripeptide gelators. Recently, rules were unravelled explaining the mechanism of gelation for these small peptides, based on their hydrophobicity and alternating chirality.[14] The peptide L-Phe-D-Ala-L-Phe was an exception to the design rules established. It was hypothesized that the hydrophobicity and steric hindrance of this molecule were not sufficient to trigger gel formation. The addition of a *p*-aminobenzoyl moiety at the N-terminus was hypothesized to solve both issues. Furthermore, this approach removed the undesirable free primary aliphatic amine (N-terminus of the peptide) which could otherwise lead to substitution of the cage aniline subcomponents thus reducing the cage stability. The newly designed gelator instead possessed an aniline substituted with an electron-deficient amide group which should not interfere with the residues from the cage.

The tripeptide (*p*-aminobenzoyl)-L-Phe-D-Ala-L-Phe-NH_2 **3.C** was thus synthesized by Ana García Fernández (Figure 3.4). Gelation occurred in CH_3CN at 30 mM after 3 h and at 50 mM within seconds after sonication of the sample. ^{19}F NMR spectrum of **3.C** revealed the presence of two TFA^- anions per molecule of **3.C**. When combined with either cage **3.1** or **3.2**, the tripeptide **3.C** formed gels in CH_3CN. For a peptide concentration of 50 mM, various cage concentrations could be used, ranging from 0.1 mM to 5 mM. The cages imparted a visible color to the gel even at concentrations as low as 0.1 mM (Figure 3.5 b and e).

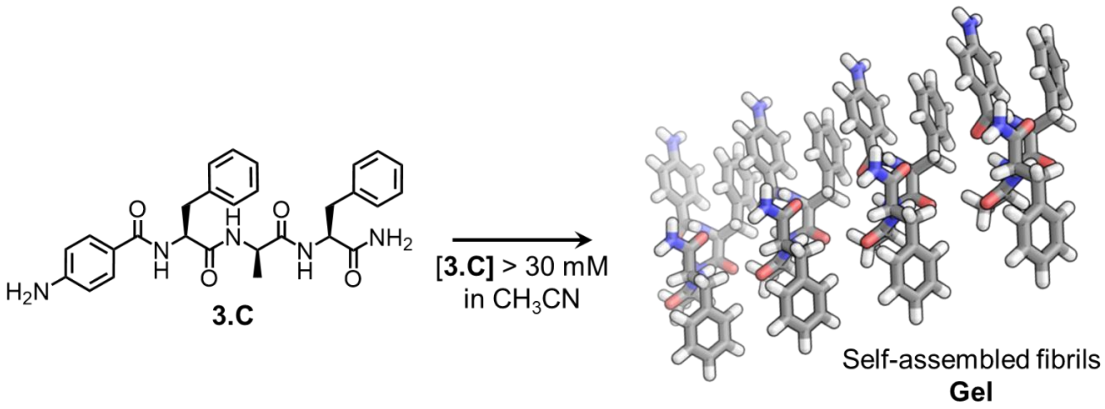


Figure 3.4 | Peptide **3.C** formed a gel at concentrations above 30 mM in CH_3CN.

Oscillatory rheometry was performed on the hybrid materials by Ana García Fernández. Time sweeps showed that the elastic modulus (G') was superior to the viscous modulus (G") for the hybrid materials **3.1⊂Gel** and **3.2⊂Gel**, confirming the gel nature of the materials (Figure 3.5 c and f). Values for the G' and G" are given in Table 3.1.

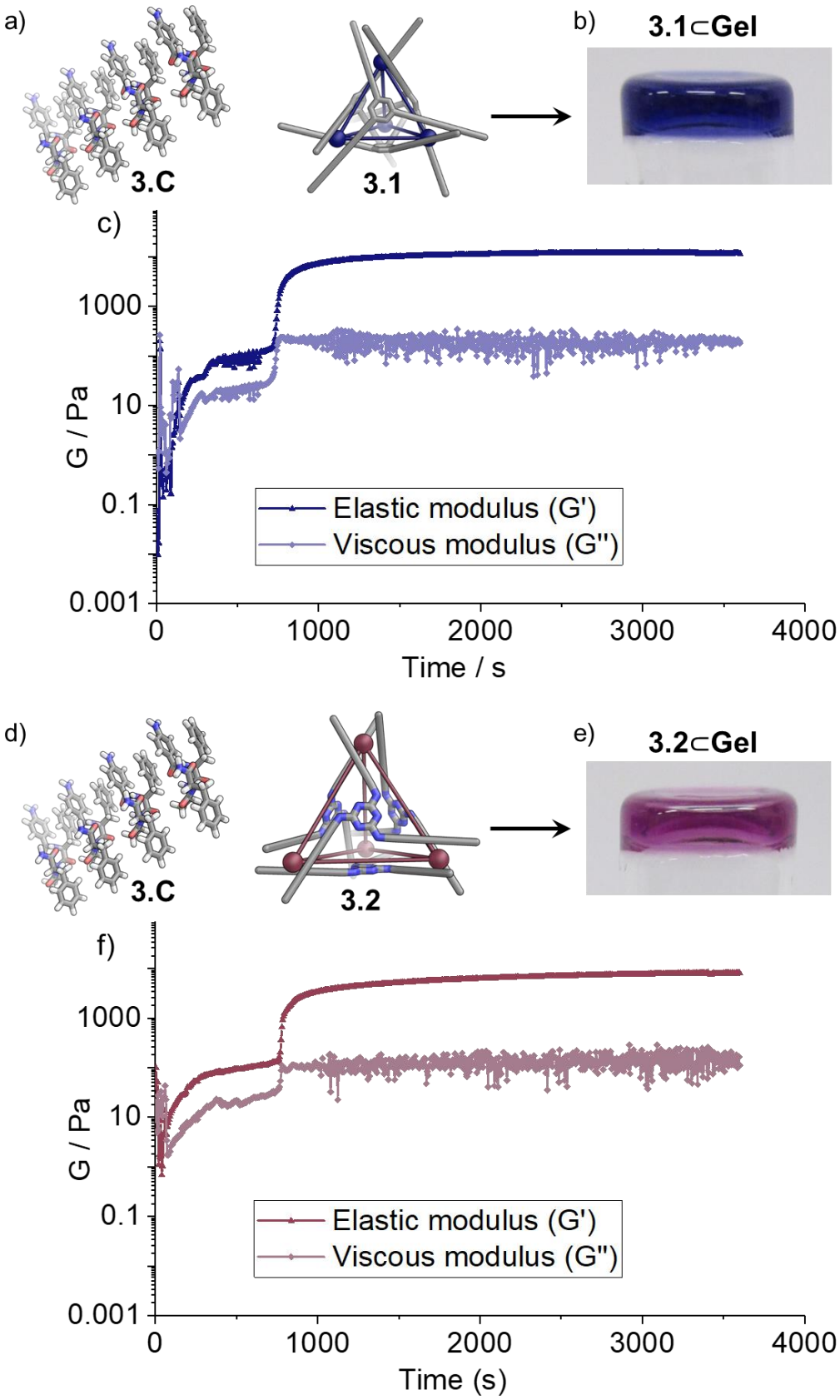


Figure 3.5 | Combining peptide **3.C** at 50 mM and either a) cage **3.1** or d) cage **3.2** at different concentrations gave the hybrid materials b) **3.1⊂Gel** and e) **3.2⊂Gel**, here pictured at 0.1 mM cage concentration. c) and f) Rheology data for the two hybrid materials at 5 mM cage concentration, showing that G' (dark blue or dark pink) is superior to G" (light blue or light pink) for both gels.

Table 3.1 | Elastic (G') and viscous (G") moduli for the gel from **3.C** and the hybrid materials ([**3.C**] = 50 mM). Average and standard deviation of three independent experiments performed using three different samples.

Sample	*G' / kPa*	*G" / kPa*
Gel from **3.C** (50 mM)	46.4 ± 4.0	1.8 ± 0.7
3.1⊂Gel ([**3.1**] = 1 mM)	18.4 ± 1.4	0.7 ± 0.3
3.2⊂Gel ([**3.2**] = 1 mM)	18.1 ± 0.5	0.9 ± 0.1
3.1⊂Gel ([**3.1**] = 5 mM)	12.3 ± 0.3	0.2 ± 0.1
3.2⊂Gel ([**3.2**] = 5 mM)	6.1 ±1.7	0.3 ± 0.1

The morphologies of the different gels were investigated by transmission electron microscopy (TEM). The images were acquired by Slavko Kralj. The peptide **3.C** gelled to give a colorless, transparent material containing a matrix of homogeneous thin fibrils of average diameter 9.8 ± 1.7 nm (Figure 3.6 a-b). When cage **3.1** or **3.2** were incorporated at a concentration of 1 mM, no significant changes to the fibril diameter were observed (8.9 ± 1.9 nm for **3.1⊂Gel** and 9.0 ± 2.1 nm for **2⊂Gel** respectively, Figure 3.6 d, e, g, h). At higher concentration (5 mM), the cages appeared to alter the fibrils' structures. The hybrid material **1⊂Gel** displayed thinner fibrils (6.5 ± 2.0 nm in diameter, Figure 3.6 j, k) than observed at lower concentrations. In contrast, **2⊂Gel** displayed thicker, shorter fibrils (11.7 ± 1.9 nm in diameter) in addition to amorphous aggregates, evidencing the disruptive effect of the cage on fibril formation (Figure 3.6 l, m).

The homogenous distribution of the cage within the material was probed by energy-dispersive X-ray (EDX) spectra which evidenced the presence of iron across all hybrid gel samples, with increasing amounts as the concentration of cage was increased (Figure 3.6 c, f, i, l, o).

Both the oscillatory rheometric data and the TEM data revealed that the cages affected the gel properties. The incorporation of cages into the gel, in addition to increasing the selectivity for guest encapsulation, was thus shown to provide a means to fine-tune gelation kinetics, stiffness and response to stress. Differences in properties were obtained when different cages were used (**3.1** or **3.2**), introducing an additional level of control over the viscoelasticity of the material.

For all further experiments, the cages were used at 5 mM to maximise their impact on the gel properties and increase the amount of guest which can be encapsulated, thus facilitating detection in the NMR experiments.

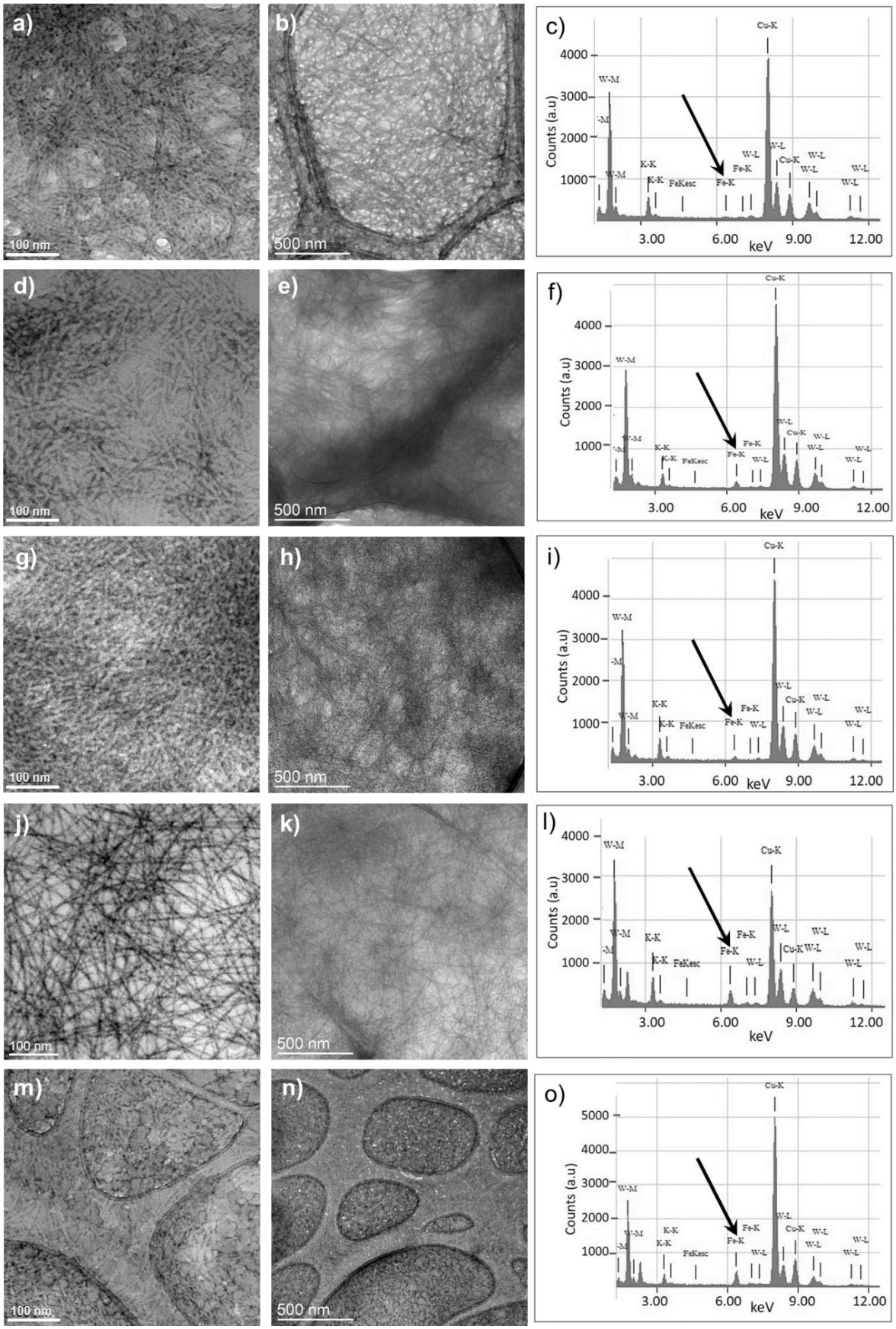


Figure 3.6 | TEM micrographs at high (left) and low (middle) magnification and corresponding EDX spectra (arrow highlighting the signal for Fe) for the gel from a-c) **3.C** and hybrid gels d-f) **3.1⊂Gel** at 1 mM, g-i) **3.2⊂Gel** at 1 mM, j-l) **3.1⊂Gel** at 5 mM and m-o) **3.2⊂Gel** at 5 mM.

The stability of cages in solution and in gels was probed by ^{1}H and ^{19}F NMR measurements. In solution, KTFA (20 eq.) was added to cages **3.1** and FA⊂**3.2** (5 mM) as peptide **3.C** contains two residual TFA^- anions (ratio obtained for hybrid gels at 50 mM peptide and 5 mM cages). Gels incorporating **3.1** or FA⊂**3.2**, were prepared and monitored over time. The degree of decomposition of each cage was inferred from the changes in the percentage of encapsulated guest. The signals for the encapsulated guest peak (either TFA^- for **3.1** or FA for **3.2**) in the ^{19}F NMR spectra were integrated against the NTf_2^- peak. The values were normalised so that the highest points were equal to 100% as only the host-guest complexes were observed at this stages. The integrals obtained for **3.1** and **3.2** in solution and **3.1⊂Gel** and **3.2⊂Gel** were plotted as a function of the time (Figure 3.7). The errors on the values of the integrals (reading error) were assumed to be 5%.

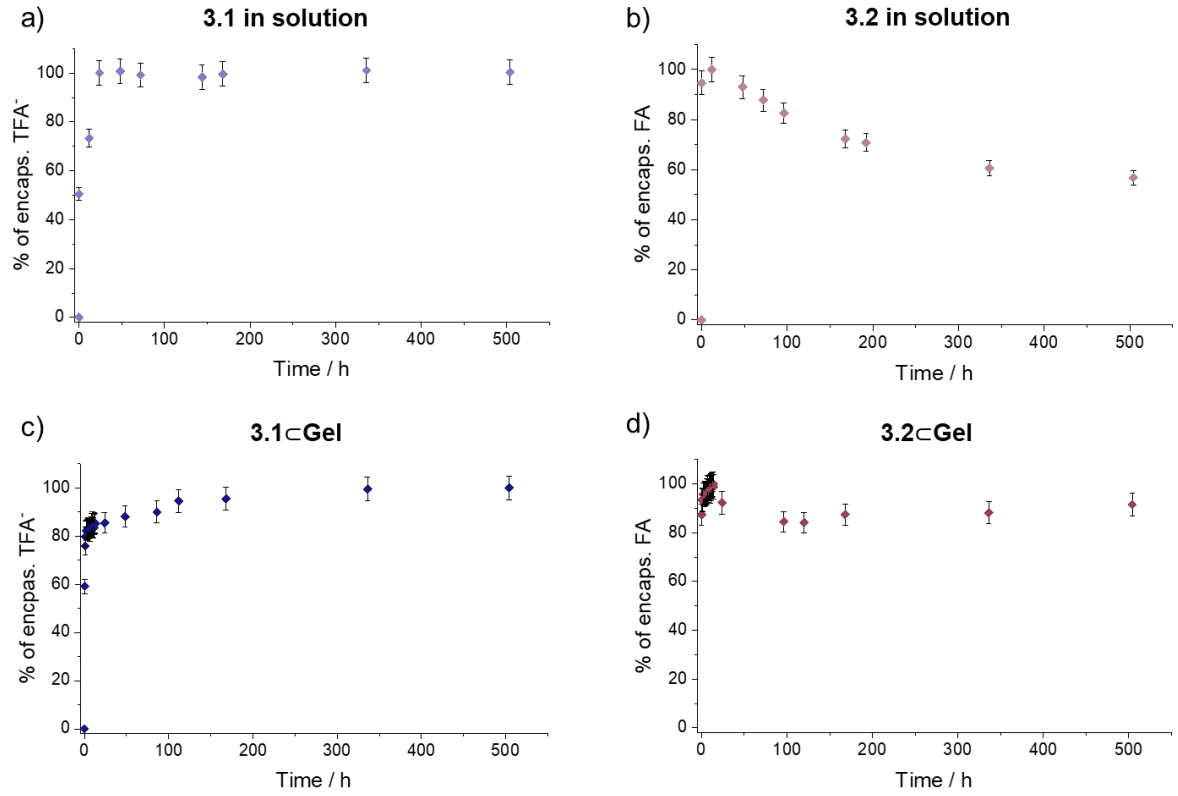


Figure 3.7 | Plots of the percentage of encapsulated guest for a) **3.1** and b) **3.2** in solution and c) **3.1⊂Gel** and d) **3.2⊂Gel**.

In both gel and solution, **3.1** encapsulated TFA^- quickly to become TFA^-⊂**3.1**. In solution, no precipitate, nor significant decrease of the intensity of the encapsulated TFA^- peak was observed over time for **3.1** (Figure 3.7 a). Similarly, the percentage of encapsulated guest did not decrease for TFA^-⊂**3.1⊂Gel** after maximum encapsulation was reached (Figure 3.7 c). Cage **3.1** was thus inferred to be stable both in solution and in the hybrid gel in the presence of

KTFA (20 eq.). Interestingly, the encapsulation of TFA^- in **3.1⊂Gel** was observed to follow slower kinetics than in solution. This was tentatively explained by competitive interactions between the TFA^- and the peptide fibrils, making the anion less available for encapsulation in **3.1**. This interaction between anions and fibrils was evidenced by broadening of the ^{19}F NMR signals corresponding to the free TFA^- in the gel, as compared to the solutions containing either **3.1** or **3.2** and KTFA. The peak corresponding to the encapsulated TFA^- was still sharp however, which was explained by the single environment found within the cage's cavity.

The percentage of FA⊂**3.2** in solution decreased steadily over long periods and a precipitate formed gradually, which was attributed to the formation of the salt FA⊂**3.2**$(TFA)_8$, expected to be insoluble in CH_3CN (Figure 3.7 b). Additionally, small amounts of free FA were observed after 2 weeks, which was attributed to a slight decomposition of **3.2**, further evidenced by the appearance of signals for subcomponent **3.B** in the ^1H NMR spectra after 2 weeks (Section 3.6, Figure 3.29). Upon formation of FA⊂**3.2⊂Gel**, small amounts of encapsulated FA were released (approximatively 7%) but no further decrease of the percentage of encapsulated guest was observed, indicative of the stability of **3.2⊂Gel** (Figure 3.7 d). While **3.2** did not appear to be stable in solution in the presence of KTFA (20 eq.), **3.2⊂Gel** was relatively stable after the initial slight decomposition. Once more, this could be explained by the TFA^- interacting with the gel fibrils, thus preventing association with **3.2** and further precipitation.

Additionally, in both **3.1⊂Gel** and **3.2⊂Gel**, the cage integrity was evidenced by the absence of NMR resonances associated with free cage subcomponents such as the aldehyde peak around 10 ppm. The stability of the cages was also probed by ESI-MS. After dilution and filtration of **3.1⊂Gel** and **3.2⊂Gel**, the spectra of the solutions obtained were acquired, showing signals for intact **3.1** and **3.2** (Figure 3.32). The spectrum for **3.1** showed a high number of peaks, some of which were assigned to the cage associated with TFA^- anions instead of NTf_2^- (see section 3.6.3).

3.3 Host-guest chemistry of cages in the gel state

The host–guest properties of **3.1⊂Gel** and **3.2⊂Gel** were studied by ^1H and ^{19}F NMR over time and compared against those of **3.1** and **3.2** in solution. Following the layering of a solution of $TBAReO_4$ or FA (1 eq.) in CD_3CN on top of **3.1⊂Gel** or **3.2⊂Gel**, the resonances assigned to free and encapsulated guests were integrated against the NTf_2^- signal in the ^{19}F NMR spectra.

Similarly, a solution of $TBAReO_4$ or FA (1 eq.) in CD_3CN was added to a solution of **3.1** or **3.2** respectively containing KTFA (20 eq.) which were previously equilibrated for 2 h, allowing for the formation of the host-guest complex $TFA^-\subset$**3.1**. As ReO_4^- could not be tracked by NMR directly, the displacement of the encapsulated TFA^- from $TFA^-\subset$**3.1** or $TFA^-\subset$**3.1**$\subset$**Gel** as a consequence of the encapsulation of ReO_4^- was monitored instead. The amount of encapsulated ReO_4^- was assumed to complement the amount of released TFA^-. The integrals obtained for **3.1** in solution and **3.1**$\subset$**Gel** were normalised so that the value for the encapsulated TFA^- is equal to 100% before the addition of TBA ReO_4 as only the host-guest complex $TFA^-\subset$**3.1** or $TFA^-\subset$**3.1**$\subset$**Gel** were observed at this stage. The amount of free FA and encapsulated FA$\subset$**3.2** were normalised so that the sum of the signals was equal to 100% at its maximum. All values obtained were plotted as a function of time (Figure 3.8 and Figure 3.9). The errors on the values of the integrals (reading error) were assumed to be 5%.

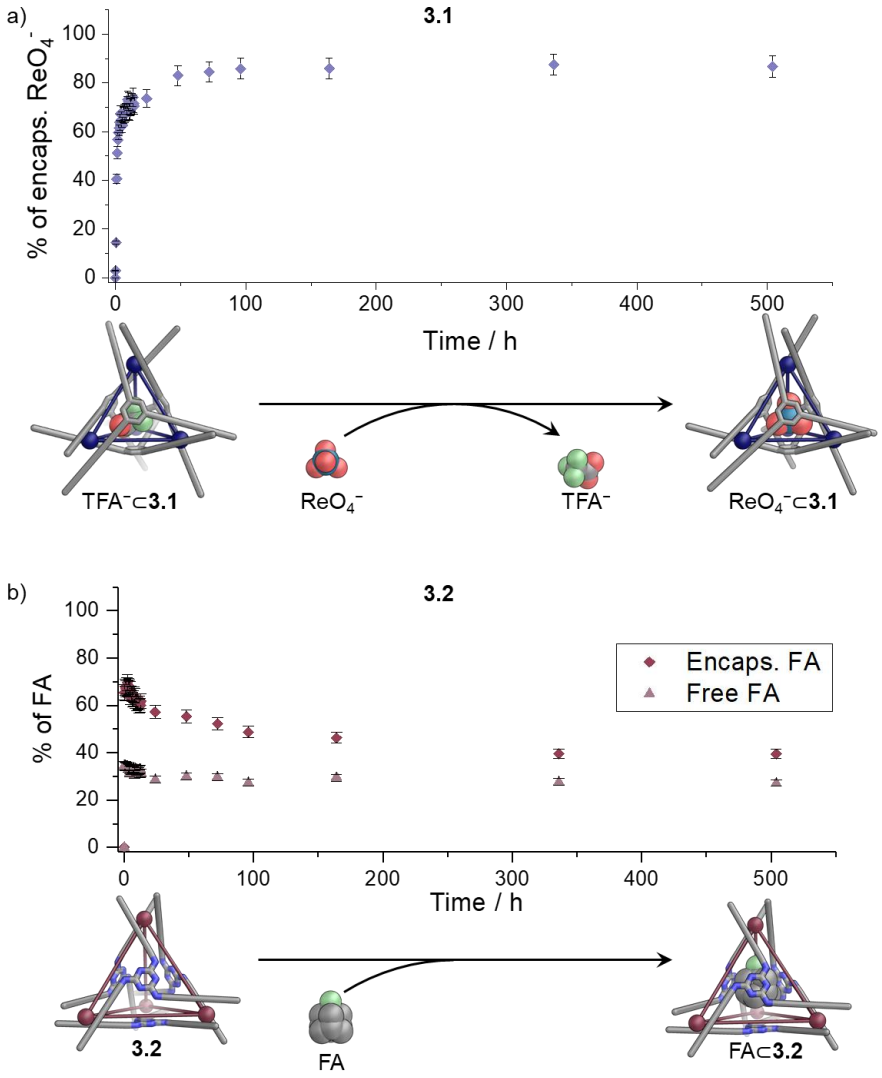


Figure 3.8 | Plots of the amount of a) encapsulated ReO_4^- for **3.1** and b) encapsulated (dark pink diamonds) and free FA (light pink triangles) for **3.2** in solution and after addition of guest (1 eq.).

The guest uptake curves in the gel state were fitted to an asymptotic exponential model (Equation 3.1) which allowed calculation of the initial rate of guest uptake and quantitative comparison between samples. In solution, the processes followed the same trends but were too fast to allow calculation of a rate of encapsulation. Similar calculations have been made in the past by Gale *et al.* to study the kinetics of anion transport through membranes.[19]

Equation 3.1: $$\%\, encaps.guest = a - b * c^{x}$$

where *x* refers to the time, *a* relates to the maximum encapsulation of a guest and *b* and *c* relate to the initial rate of encapsulation by the following formulae (Equation 3.2):

Equation 3.2: $$k_{ini} = -b * ln(c)$$

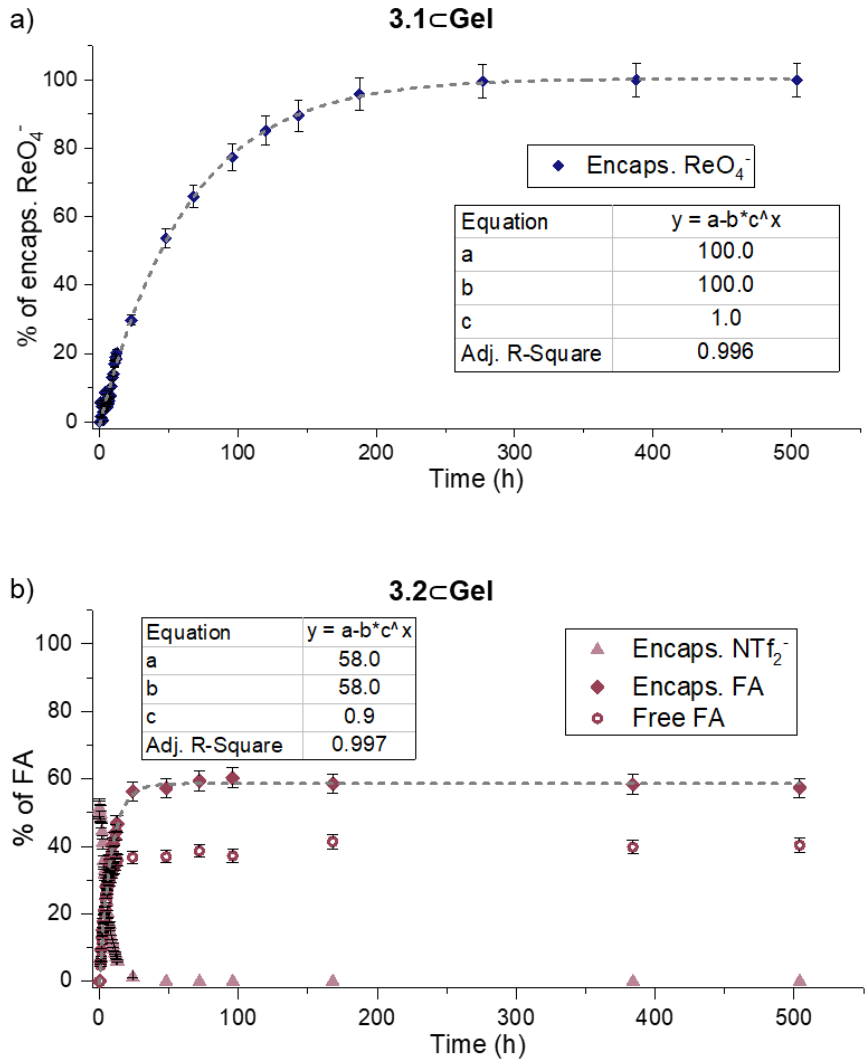


Figure 3.9 | Plots of the amount of a) encapsulated ReO_4^- for **3.1⊂Gel** and b) encapsulated NTf_2^- (light pink triangles), encapsulated FA (dark pink diamonds) and free FA (dark pink circles) for **3.2⊂Gel** after addition of guest (1 eq.). The grey dotted lines represent the fitting from the parameters given in the insert table.

In solution, the maximum encapsulation of ReO_4^- in **3.1** (86% ± 4) was reached after four days (Figure 3.8 a) whereas the maximum encapsulation of FA in **3.2** (69% ± 3) was reached within 2 h of guest addition (Figure 3.8 b). The subsequent decrease in FA⊂**3.2** followed the same trend previously observed for the stability tests in solution and was attributed to precipitation of **3.2**. The complete encapsulation of ReO_4^- in **3.1⊂Gel** was reached after 12 days (Figure 3.9 a). For **3.2⊂Gel**, the signal for the encapsulated NTf_2^- dropped rapidly, with no trace of encapsulation for this compound after 48 h and the maximum encapsulation of FA (59% ± 3) was reached after 72 h (Figure 3.9 a). More ReO_4^- was encapsulated in **3.1⊂Gel** relative to **3.1** in solution, which was hypothesized to result from the binding of TFA^- anions by the gel fibrils, thus reducing the TFA^- availability for competitive encapsulation.

Comparison of the rates of guest uptake between solution and gel state clearly highlighted the slower kinetics in the gels (Figure 3.10). It was thus inferred that the rate of guest uptake by cages in the gel only partially resulted from the rate of guest encapsulation and that other effects were involved as well. Based on the striking difference observed between **3.1⊂Gel** and **3.2⊂Gel** and the known difference in their properties on the material level, the rate of guest uptake was hypothesized to be dependent of the kinetics of diffusion through the sample in addition to that of guest encapsulation.

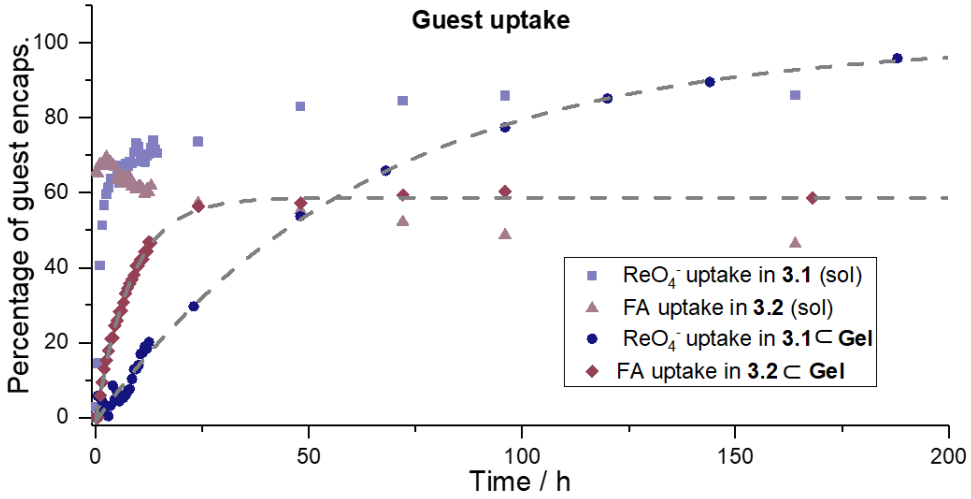


Figure 3.10 | Overlap of the plots showing the amount of encapsulated guest in **3.1** in solution (light blue squares), **3.2** in solution (light pink triangles), **3.1⊂Gel** (dark blue circles) and **3.1⊂Gel** (dark pink diamonds) after addition of guest (1 eq.). The grey dotted lines represent the fitted model.

The diffusion of $TBAReO_4$ was gauged by monitoring the signals for TBA^+ in the ^1H NMR spectra and the diffusion of FA was gauged by monitoring the signals for both free and encapsulated FA in the ^{19}F NMR spectra (Figure 3.11).

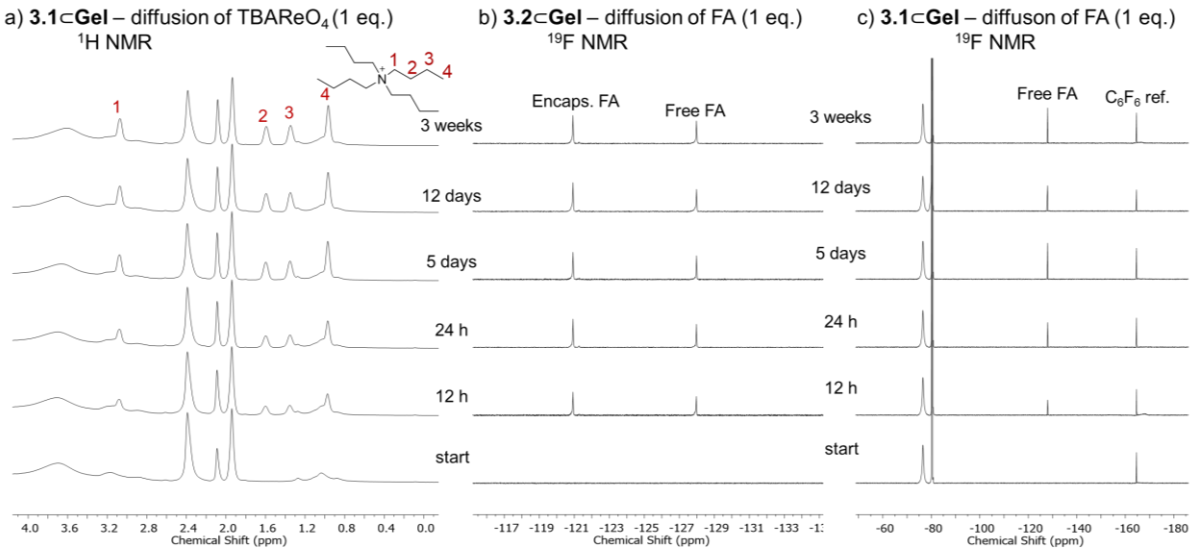


Figure 3.11 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) of **3.1⊂Gel** after addition of $TBAReO_4$ (1 eq.) showing the signals for TBA^+ over time and ^{19}F NMR spectra (471 MHz, 298K, CD_3CN) of b) **3.2⊂Gel** and c) **3.1⊂Gel** after addition of FA (1 eq.). Time elapsed from guest addition is given on the spectra.

The diffusion curves were fitted to the same asymptotic exponential model as previously (Equation 3.1) and the initial rates of diffusion were calculated. In **3.1⊂Gel**, the rates of diffusion for ReO_4^- (5.44 ± 0.20 % h^{-1}) and FA (6.30 ± 0.22 % h^{-1}) were similar whereas the rate of diffusion of FA in **3.2⊂Gel** (14.61 ± 0.28 % h^{-1}) was faster than in **3.1⊂Gel** (Figure 3.12). These observations suggest that the rate of diffusion of small molecules within the hybrid gels is mainly influenced by the structure of the gel network, rather than by the chemical structure of the diffusing species. Indeed, the higher storage modulus G' of **3.1⊂Gel** relative to **3.2⊂Gel** indicates that the first gel is more "robust" and hence harder to diffuse through. Furthermore, it was inferred that the difference in nanostructures between gels, such as thinner fibrils with higher surface area, lead to greater physisorption of guest molecules and therefore slower diffusion.

Overall, a five-fold difference between the initial rate of guest uptake for **3.1⊂Gel** and **3.2⊂Gel** (1.60 ± 0.05 % h^{-1} and 7.23 ± 0.15 % h^{-1} respectively) was obtained due to the different properties of the hybrid materials (Figure 3.13). The difference resulted primarily from the microscopic structure of the material and the kinetic of guest encapsulation within the cage to a lesser extent. Temporal control of the encapsulation of small guest molecules could thus be achieved for this system.

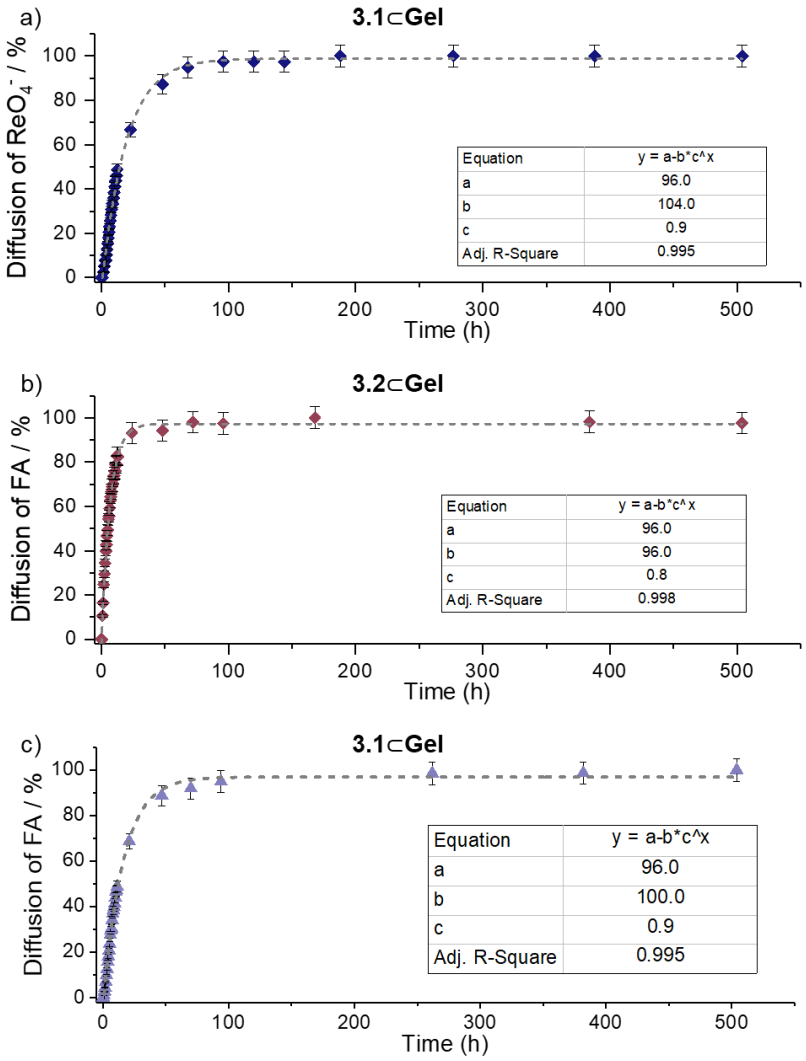


Figure 3.12 | Plots of the diffusion of a) $TBAReO_4$ in **3.1⊂Gel**, b) FA in **3.2⊂Gel** and c) FA in **3.1⊂Gel**. The grey dotted lines represent the fitting from the parameters given in the insert tables.

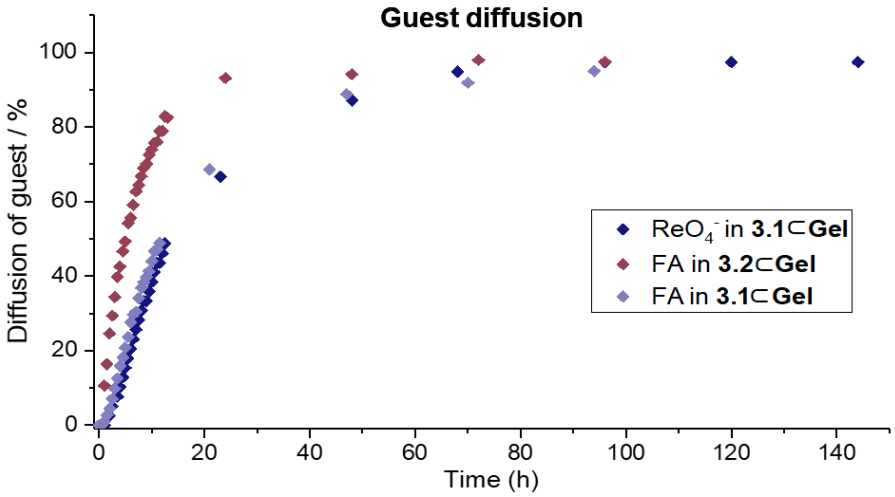


Figure 3.13 | Overlap of the plots showing the diffusion of $TBAReO_4$ in **3.1⊂Gel** (dark blue), FA in **3.2⊂Gel** (dark pink) and FA in **3.1⊂Gel** (light blue) highlighting the similarity in diffusion kinetics between $TBAReO_4$ and FA in **3.1⊂Gel** and the difference between diffusion of FA in **3.1⊂Gel** and **3.2⊂Gel.**

Finally, the diffusion of cages through the hybrid gels was investigated. A solution of TFA^-⊂**3.1** or FA⊂**3.2** (1 mM, 50 µL) was layered on top of hybrid gels **3.2⊂Gel** or **3.1⊂Gel** respectively. The diffusion was monitored *via* the integrals of the encapsulated TFA^- peak for **3.2⊂Gel** and the encapsulated FA for **3.1⊂Gel** in the ^{19}F NMR spectra.

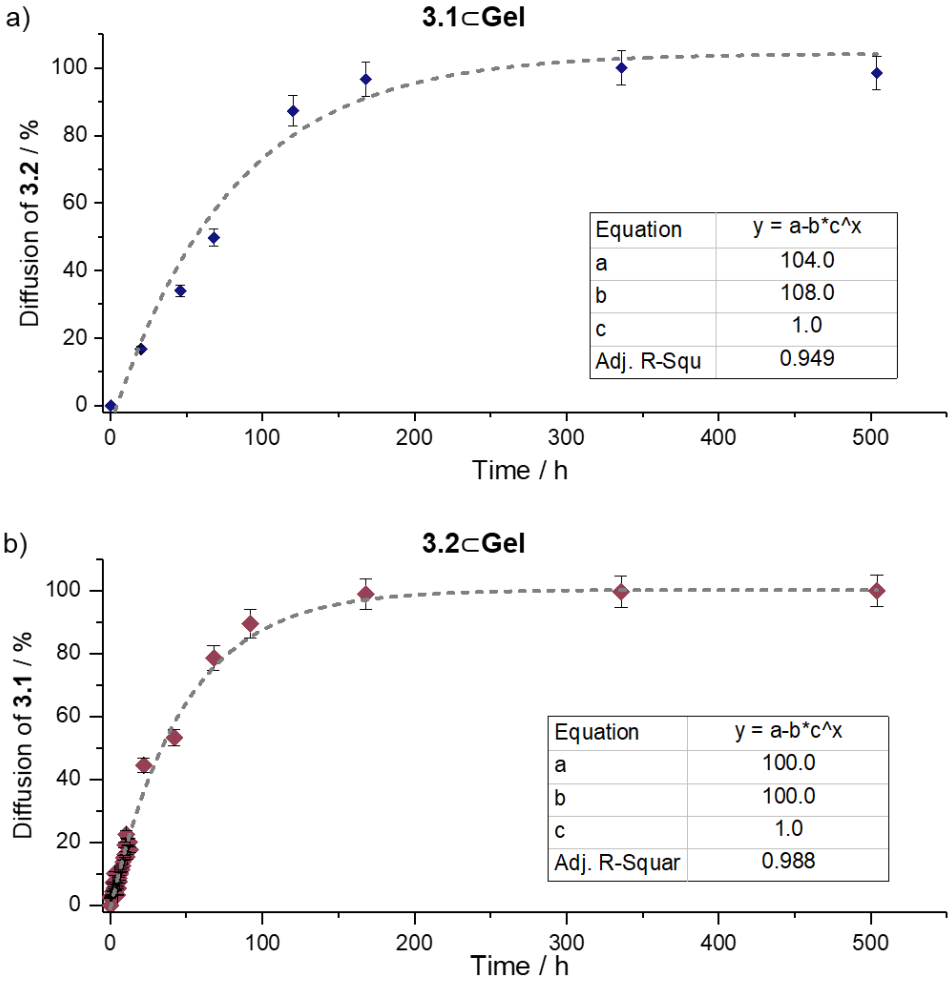


Figure 3.14 | Plots of the diffusion of a) FA⊂**3.2** in **3.1⊂Gel** and b) TFA^-⊂**3.1** in **3.2⊂Gel**. The grey dotted lines represent the fitting from the parameters given in the insert tables.

As expected, the diffusion of the cages was slower than that of the smaller molecular guests (Figure 3.14). The initial rate of diffusion of **3.1** was seven times slower compared to FA in **3.2⊂Gel** (Figure 3.15, solid vs. dash-dot line). The initial rate of diffusion of **3.2** was four times slower compared to ReO_4^- in **3.1⊂Gel** (Figure 3.15, dashed vs. dotted line). This implies that the cages remain localized in the gel matrix on a timescale that allowed the smaller guests to diffuse through the sample. The hybrid gel system was therefore suitable for cage segregation in separate layers and selective guest encapsulation. All values obtained for guest uptake and diffusion are summarised in Table 3.2.

Table 3.2 | Summary of the initial rate of uptake and diffusion for ReO_4^-, FA, **3.1** and **3.2** in **3.1**⊂Gel and **3.2**⊂Gel derived from kinetic ^1H and ^{19}F NMR experiments.

	Initial rate of uptake (% h^{-1})		Initial rate of diffusion (% h^{-1})			
	ReO_4^-	FA	ReO_4^-	FA	**3.1**	**3.2**
3.1⊂**Gel**	1.60 ± 0.05		5.44 ± 0.20	6.30 ± 0.22		1.31 ± 0.20
3.2⊂**Gel**		7.23 ± 0.15		14.61 ± 0.28	2.11 ± 0.11	

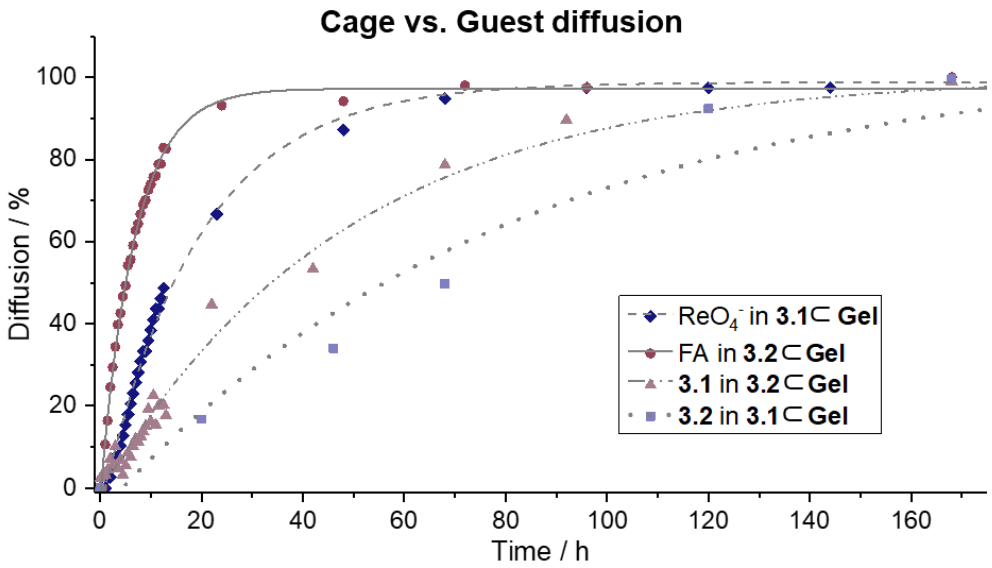


Figure 3.15 | Overlap of the plots showing the diffusion of $TBAReO_4$ in **3.1**⊂**Gel** (dark blue diamonds, dashed grey line), FA in **3.2**⊂**Gel** (dark pink circles, solid grey line), FA⊂**3.2** in **3.1**⊂**Gel** (light blue squares, grey dotted line) and TFA^-⊂**3.1** in **3.2**⊂**Gel** (light pink triangles, dash-dot grey line).

3.4 Chemical separation within a tri-layered system

With guest separation in mind, a tri-layered system composed of **3.1**⊂**Gel** and **3.2**⊂**Gel**, separated by a buffer layer of peptide gel devoid of cages (**3.C** only) was designed (Figure 3.16). A solution of cage **3.1** in CD_3CN (5 mM, 0.12 mL) was combined with **3.C** (4.4 mg, 6.0 µmol, 10 eq.). The solution was transferred in a 5 mm NMR tube and was sonicated for 10 min to promote the formation of the gel. After 2 h of equilibration, a solution of **3.C** in CD_3CN (50 mM, 0.09 mL) was layered on top of the gel containing **3.1**. The sample was sonicated for 10 min to promote the formation of the gel (Figure 3.16 a). This layer served as a buffer between the two cages. A solution of cage **3.2** in CD_3CN (5 mM, 0.12 mL) was combined with **3.C** (4.4 mg, 6.0 µmol, 10 eq.). The solution was layered on top of the sample and was sonicated for 10 min to promote the formation of the gel (Figure 3.16 b).

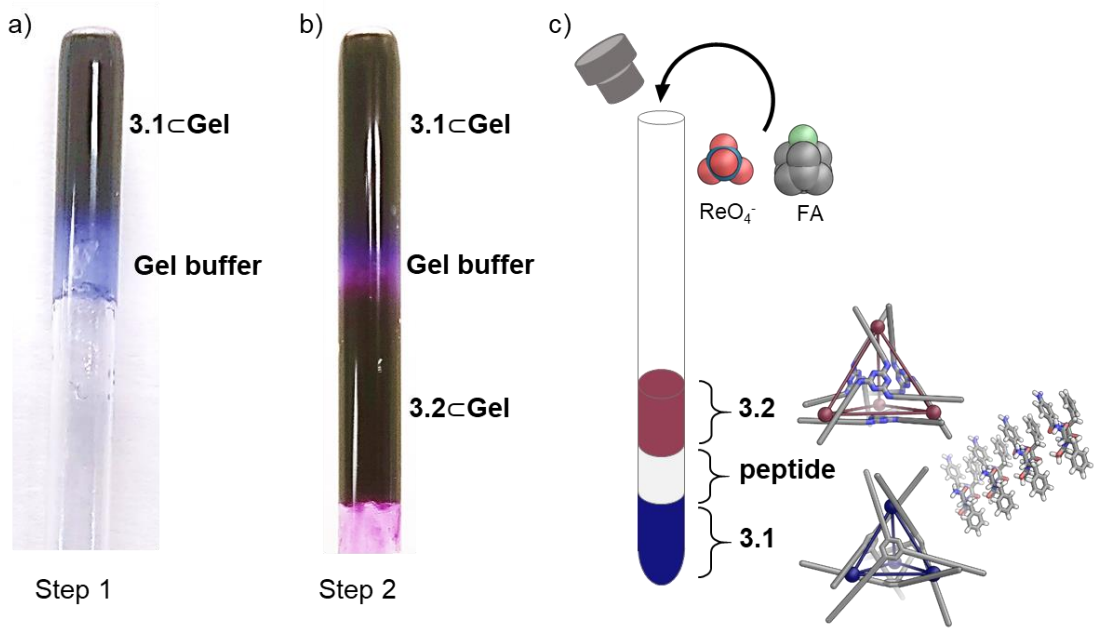


Figure 3.16 | Composition of the tri-layered system. a) Photo of **3.1⊂Gel** (dark blue) and the buffer (light blue). b) Photo of the complete tri-layered system with **3.1⊂Gel** (dark blue), buffer (light blue/purple) and **3.2⊂Gel** (dark purple). c) Schematic representation of the tri-layered system.

The sample was studied by slice-selective ^{1}H NMR.[20-21] ^{1}H NMR 2D mapping was done by acquiring 20 slices of approximately 1 mm thickness to obtain a high resolution spatial representation of the sample. The 'map' of the sample produced allowed the selection of the slices with the best homogeneity and signal intensity for the monitoring of the process.

The presence of **3.1** and **3.2** in their respective layers was confirmed by spatial mapping of the sample (Figure 3.17). The grey regions indicate the spatial partitioning of the proton signals corresponding to cages **3.1** and **3.2** within the gel matrix. ^{1}H NMR signals assigned to **3.1** (blue spectrum) were only found in layer 1, and signals for **3.2** (purple spectrum) were only found in layer 2. The observation of color leaching into the buffer layer was attributed to the strong visible absorbance of MOCs even at low concentrations (Figure 3.5 b and e). ^{1}H NMR signals of the peptide only were detected there, indicating that the MOCs remained segregated mostly in their original layers and were not present at concentrations above the NMR detection threshold in the buffer layer.

After the tri-layered gel formation, a solution of $TBAReO_4$ and FA in CD_3CN (1 eq., 0.02 mL) was added on top of the sample, and the encapsulation of the guests in the different gel layers was followed by slice-selective ^{19}F NMR (Figure 3.18 and Figure 3.19). Two regions of the ^{19}F NMR (−135 to − 115 ppm and −88 to − 68 ppm) were monitored over time for layer 1 and layer 2.

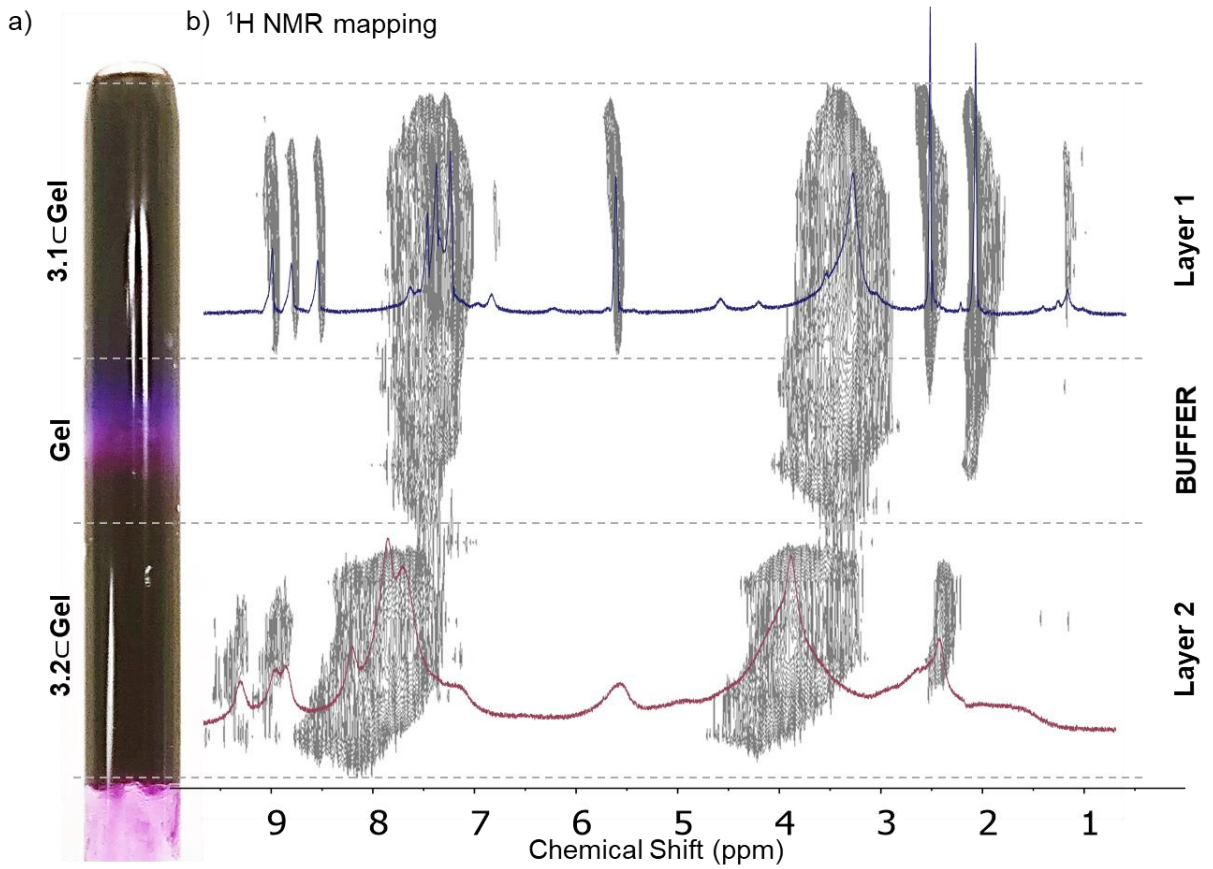


Figure 3.17 | a) Photo of the sample and b) 2D mapping of the ^{1}H NMR spectra for the three-layered gel, showing the presence of **3.1** in layer 1 (**3.1⊂Gel**), the peptide alone in the buffer gel layer, and **3.2** in layer 2 (**3.2⊂Gel**). The ^{1}H NMR spectra of **3.1⊂Gel** (blue) and **3.2⊂Gel** (pink) obtained by slice-selective NMR are overlapped for clarity.

The signal corresponding to encapsulated FA was observed in layer 2, with maximum encapsulation reached after 12 h (Figure 3.18 b). Free FA was observed in both layer 1 (Figure 3.19 b) and layer 2 (Figure 3.18 b), which was explained by the weaker binding of this guest in **3.2**. It is important to note that no comparisons between the intensity of the signals observed in both layers could be made as the performance of the detection greatly varied from one layer to another, as evidenced by the higher signal/noise ratio in layer 2 compared to layer 1. Layer 2 additionally presented signals for FA⊂**3.2**, which were only observed at later stages (t > 24h) in layer 1 due to leaching of FA⊂**3.2**. Despite the lack of absolute separation of FA between layers, an enrichment of this compound was achieved in layer 2.

The disappearance of the encapsulated TFA^{-} peak, indicative of ReO_4^{-} encapsulation was observed in layer 1 (Figure 3.19 a). Approximately 80% of ReO_4^{-} was observed to be encapsulated in **3.1** (by integration against the triflimide peak) after 62 h. The remaining 20% was inferred to be entrapped in the gel and spread evenly through the sample. Even though

complete segregation of the ReO_4^- in layer 1 was not achieved, an enrichment of up to 9:1 ratio in layer 1 compared to layer 2 was obtained. The potential of this system for chemical separation was thus demonstrated.

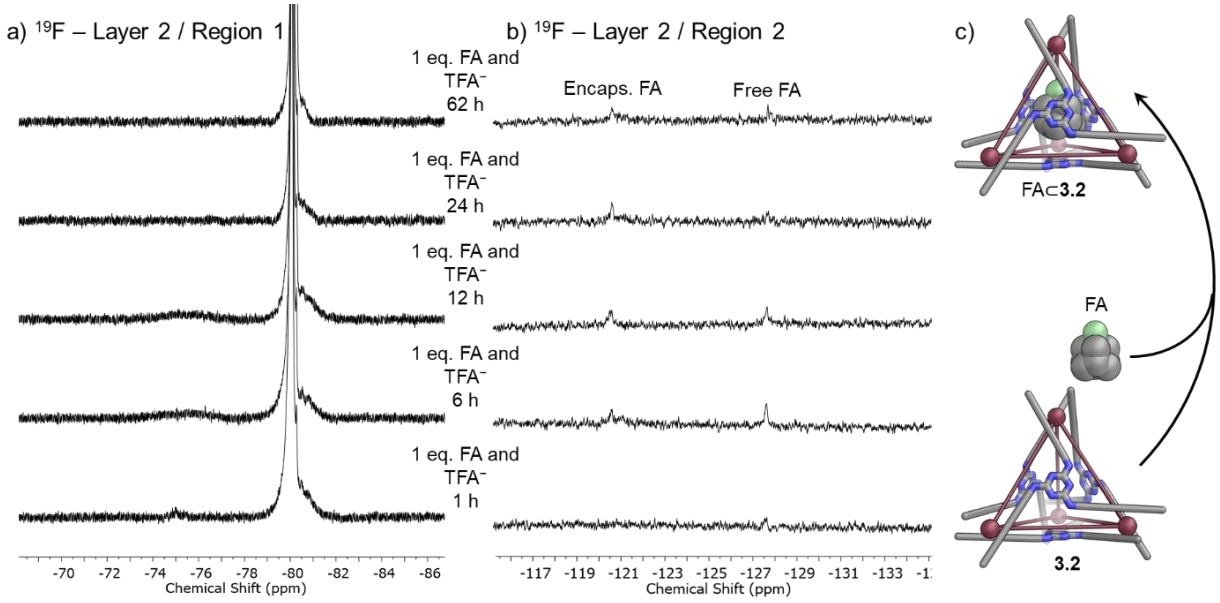


Figure 3.18 | a and b) ^{19}F NMR spectra (471 MHz, 298 K, CD_3CN) of layer 2 containing **3.2⊂Gel** in the two regions where the guests signals are detected after addition of a mixture of FA and ReO_4^- (1eq.). Time elapsed from guest addition is given on the spectra. c) Cartoon representation of the chemical process.

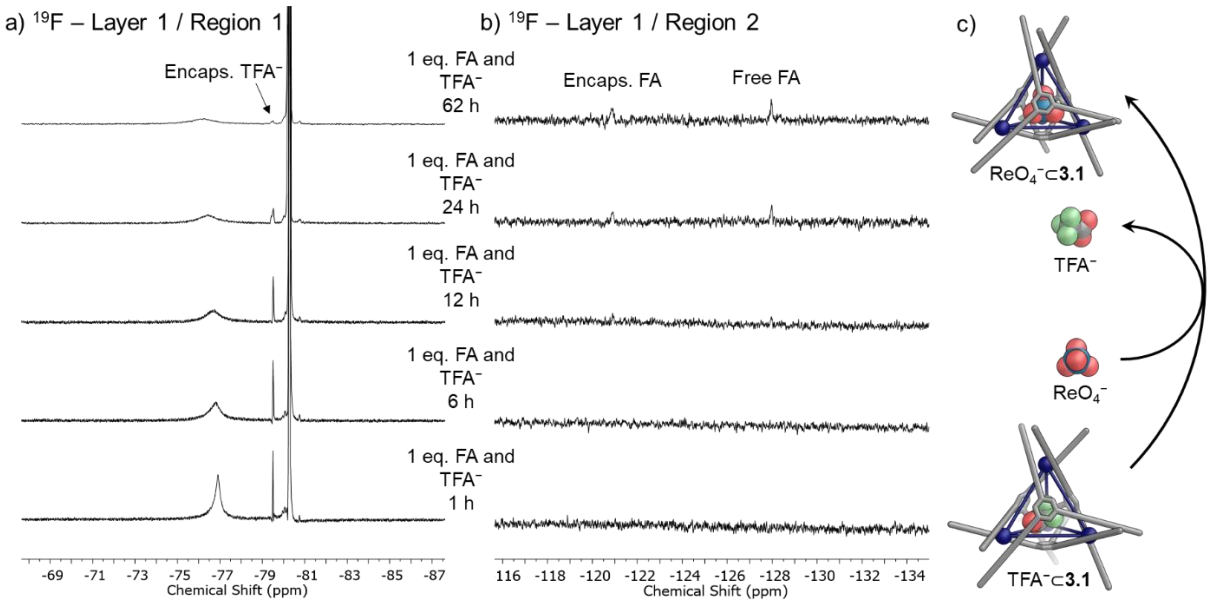


Figure 3.19 | a and b) ^{19}F NMR spectra (471 MHz, 298 K, CD_3CN) of layer 1 containing **3.1⊂Gel** in the two regions where the guests signals are detected after addition of a mixture of FA and ReO_4^- (1eq.). Time elapsed from guest addition is given on the spectra. c) Cartoon representation of the chemical process.

3.5 Conclusions and future work

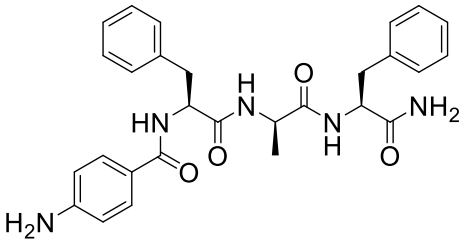
In conclusion, the formation of hybrid cage⊂gel nanostructured materials made of a self-assembled tripeptide and one of two distinct $Fe^{II}_4L_4$ cages was achieved. Once embedded in the gel, the cavities of the cages were still available for guest uptake and either ReO_4^- or FA could be encapsulated. The gel nanostructure was shown to be influenced by the presence of the cages, thus influencing the rate of diffusion of different compounds through the material based on their size. This provided new means for the control of diffusion kinetics and, as a consequence, the uptake of small molecules. These hybrid gels allowed for the creation of spatially separated, chemically-distinct phases within one-solvent systems. After embedding two different cages in distinct areas of the gel, the supramolecular assemblies allowed for chemical segregation of molecules from a mixture by means of guest encapsulation.

The chemical platform established in this chapter could be extended in the future to create biocompatible materials. Recently, Nitschke *et al.* reported a blueprint for the formation of water soluble self-assembled cages.[22] Combined with peptide LMWGs that can assemble in water, these water soluble cages could generate new, easily accessible materials able to perform guest uptake and release under biological conditions. Potential uses in drug delivery could thus be envisioned.

3.6 Experimental section

3.6.1 Synthesis of 3.C, 3.1 and 3.2

Peptide p-aminobenzoyl-L-Phe-D-Ala-L-Phe-NH₂ 3.C



p-aminobenzoyl-L-Phe-D-Ala-L-Phe-NH_2 **3.C** was synthesized by AMG following Fmoc-based solid phase peptide synthesis under a dry and inert atmosphere. The resin 2-chlorotrytil chloride (10 g) was swollen in DCM (40 mL). $SOCl_2$ (1 mL) was added and the reaction was shaken under an argon for 1 h. The resin was washed with DMF (2 × 30 mL) and DCM (2 × 30 mL). A solution of Fmoc-Rink amide linker (5.4 g, 10 mmol) and DIPEA (9 mL) in

DMF/DCM (2:1) was added to the resin, and the mixture was stirred for 1.5 h. MeOH (4 mL) was added and the reaction mixture was shaken for 5 min and then washed with DMF (3 × 30 mL) and dicloromethane (3 × 30 mL). Piperidine 20% in DMF (10 mL) was added to the reactor and was stirred at room temperature (2 × 10 minutes) for deprotection. The reaction mixture was washed with DMF and DCM. A mixture of Fmoc-L-Phe-OH (7.3 g, 18.8 mmol), HBTU (5.7 g, 15.0 mmol), HOAt (2.0 g, 15.0 mmol) and DIPEA in DMF (1M, 15 mL) was sonicated until clear, and was added to the reactor. The reactor was shaken at room temperature for 1.5 h. The resin was washed and deprotected as in the previous step. The coupling and deprotection of the amino acids D-Ala and L-Phe were done the same way as the first coupling, by using Fmoc-D-Ala-OH (5.9 g, 18.8 mmol) for the second coupling and Fmoc-L-Phe-OH (7.3 g, 18.8 mmol) for the third coupling. The introduction of *p*-aminobenzoyl motif, was performed under the same coupling conditions by using Boc-*p*-aminobenzoic acid instead of an Fmoc-amino acid. The peptide was cleaved from the resin by shaking 2 h in the presence of a solution of TFA/DCM/H_2O/triisopropylsilane (47.5/47.5/2.5/2.5, 100 mL). The solution was drained from the reactor, and the solvent was evaporated under air flow. The remaining oil was dissolved in a mixture of CH_3CN/H_2O containing 0.05% of TFA, and purified by reverse-phase HPLC (Agilent Technologies, Santa Clara, CA, USA). The HPLC Agilent 1260 Infinity system was equipped with a preparative gradient pump (1311B), semi-preparative C-18 column (Kinetex, 5 microns, 100 Å, 250 mm x 10 mm, Phenomenex, Torrance, CA, USA), autosampler (G1329B), and Photodiode Array detector (G1315C). The following HPLC method was used for the purification of the peptide: t = 0–2 min, 25% CH_3CN; t = 14 min, 80% CH_3CN; t = 16 min, 95% CH_3CN; t = 17 min, 95% CH_3CN. The compound was then freeze-dried to yield the corresponding peptide as a white powder. ^1H and ^{19}F NMR were used to assess the amount of residual TFA^- per peptide. Integration of ^1H and ^{19}F signals against a reference capillary of 1,3,5-trifluorobenzene showed the presence of 2 eq. of TFA^- per equivalent of peptide.

^{1}H NMR (400 MHz, 298K, DMSO-d_6) δ 8.19 (d, J = 7.5 Hz, 1H, H_d), 8.10 (dd, J = 8.2, 1.8 Hz, 2H, $H_{j,m}$), 7.55 (d, J = 8.8 Hz, 2H, H_c), 7.36 (bs, 2H, H_s), 7.29 – 7.07 (m, 10H, $H_{g,h,i,p,q,r}$), 6.56 (d, J = 8.8 Hz, 2H, H_b), 4.75 (bs, 2H, H_a), 4.52 (ddd, J = 9.1, 7.5, 5.9 Hz, 1H, H_e), 4.35 (ddd, J = 10.6, 8.8, 4.1 Hz, 1H, H_n), 4.16 (q, J = 7.3 Hz, 1H, H_k), 3.11 – 2.91 (m, 3H, $H_{f,o}$), 2.75 (dd, J = 13.7, 10.6 Hz, 1H, H_o), 0.91 (d, J = 7.0 Hz, 3H, H_l). **^{13}C NMR** (100 MHz, 298K, DMSO-d_6) δ 173.4, 172.1, 171.8, 167.7, 151.1, 138.1, 137.5, 129.3, 129.2, 129.1, 128.4, 128.2, 127.9, 126.6, 126.4, 121.6, 117.3, 113.5, 56.4, 55.0, 48.9, 37.0, 36.7, 16.2. **ESI-MS**: m/z

calculated for M = 501.2, observed positive mode $[M + H]^+$ = 502.2 and negative mode $[M - H]^-$ = 500.2. **HRMS:** *m/z* calculated for [**3.C** + H^+] = 502.2449, observed = 502.2398.

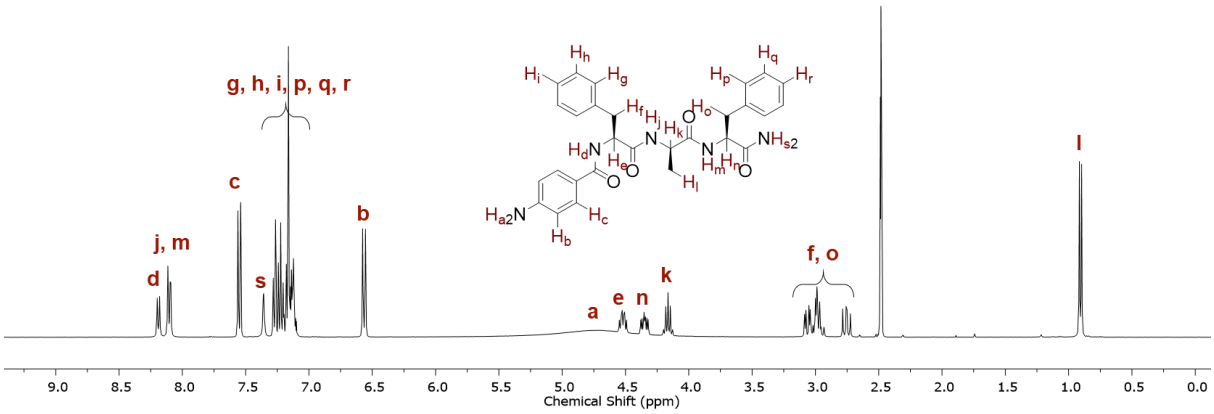


Figure 3.20 | ^{1}H NMR spectrum (400 MHz, 298 K, DMSO-d_6) of **3.C**

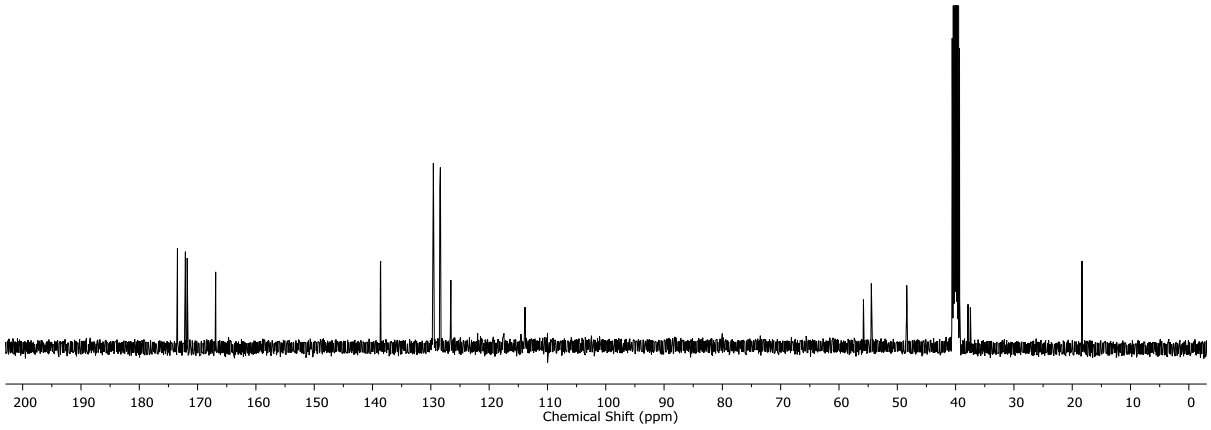


Figure 3.21 | ^{13}C NMR spectrum (101 MHz, 298 K, DMSO-d_6) of **3.C**.

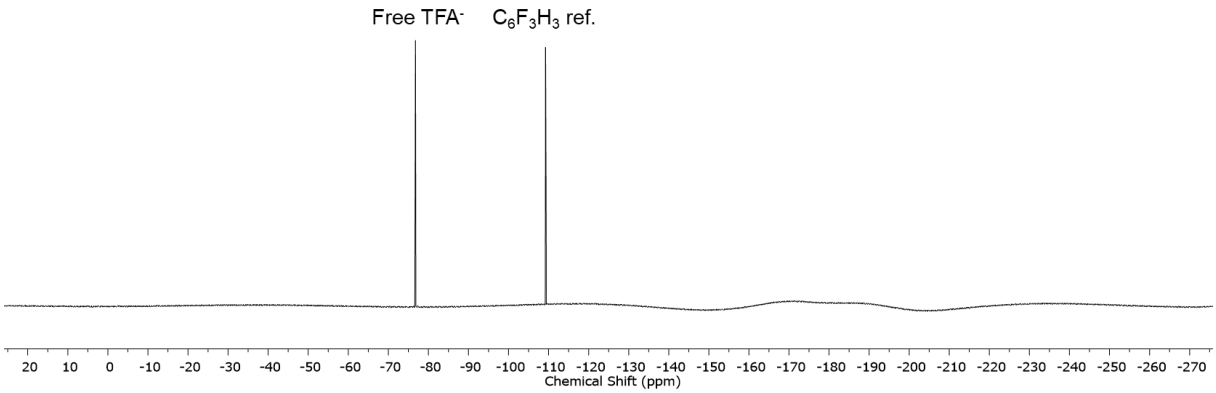


Figure 3.22 | ^{19}F NMR spectrum (471 MHz, 298 K, CD_3CN) of **3.C**.

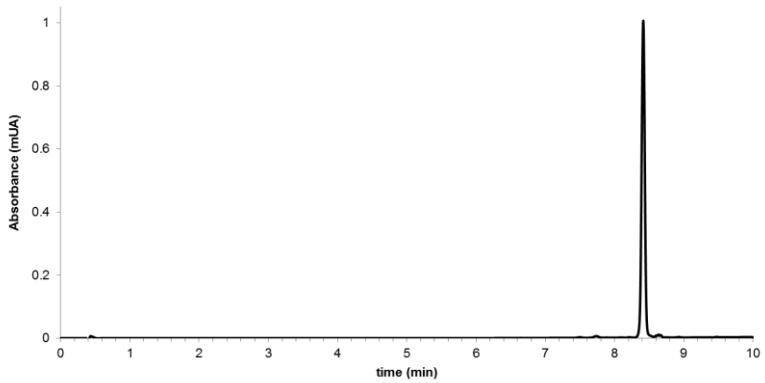
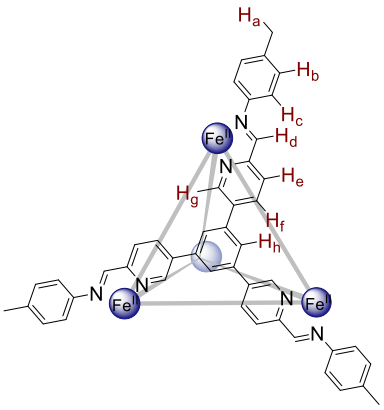


Figure 3.23 | LC trace of **3.C** (λ = 254 nm). Method: t = 0, 95% water (+0.1% CF_3COOH) and 5% CH_3CN (+0.1% CF_3COOH); t = 10 min, 5% water (+0.1% CF_3COOH) and 95 % CH_3CN (+0.1% CF_3COOH). Flow: 0.5 ml/min. (t_R= 8.4 min).

Cage 3.1



5,5',5"-(benzene-1,3,5-triyl)tripicolinaldehyde **3.A** (40.0 mg, 0.1018 mmol, 1.0 eq.) and *p*-toluidine (32.7 mg, 0.3054 mmol, 3.0 eq.) were dissolved in CH_3CN (10 mL) in a sealed 25 mL round bottom flask. The solution was degassed with N_2 for 10 min after which $Fe(NTf_2)_2$ (70.0 mg, 0.1018 mmol, 1.0 eq.) was added. The solution was degassed for an additional 10 min. The solution was heated at 50 °C for 18 h. The dark blue solution was then cooled and concentrated under a flow of nitrogen. Addition of Et_2O (80 mL) precipitated the compound as a dark blue solid. The solid was separated by centrifugation and washed with Et_2O (2 × 80 mL). The solid was then dried under a flow of N_2.

^{1}H NMR (500 MHz, 298K, CD_3CN) δ 8.95 (s, 12H, H_d), 8.81 (d, *J* = 8.3, 12H, H_e), 8.55 (dd, *J* = 8.3, 2.0 Hz, 12H, H_f), 7.43 (s, 12H, H_h), 7.14 (d, *J* = 8.1 Hz, 24H, H_b), 7.11 (d, *J* = 2.0 Hz, 12H, H_g), 5.51 (d, *J* = 8.0 Hz, 24H, H_c), 2.41 (s, 36H, H_a). **^{19}F NMR** (471 MHz, 298K, CD_3CN) δ –80.45. **LR-ESI-MS** [charge, calculated mass]: *m/z* = 1422.2 [**3.1**$(NTf_2)_5^{3+}$, 1422.4], 996.7 [**3.1**$(NTf_2)_4^{4+}$, 996.8], 741.3 [**3.1**$(NTf_2)_3^{5+}$, 741.4], 571.0 [**3.1**$(NTf_2)_2^{6+}$, 571.2], 449.4 [**3.1**$(NTf_2)^{7+}$, 449.5], 358.2 [**3.1**$^{8+}$, 358.3].

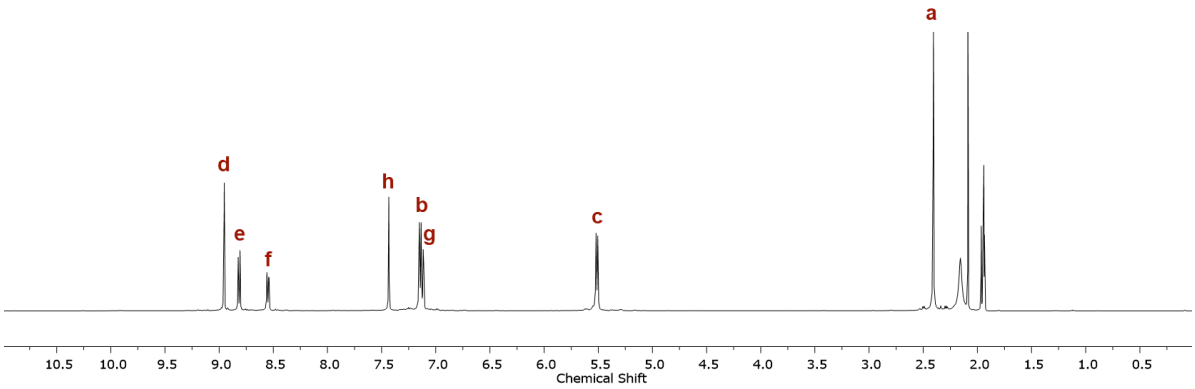


Figure 3.24 | ^{1}H NMR spectrum (500 MHz, 298 K, CD_3CN) of **3.1**.

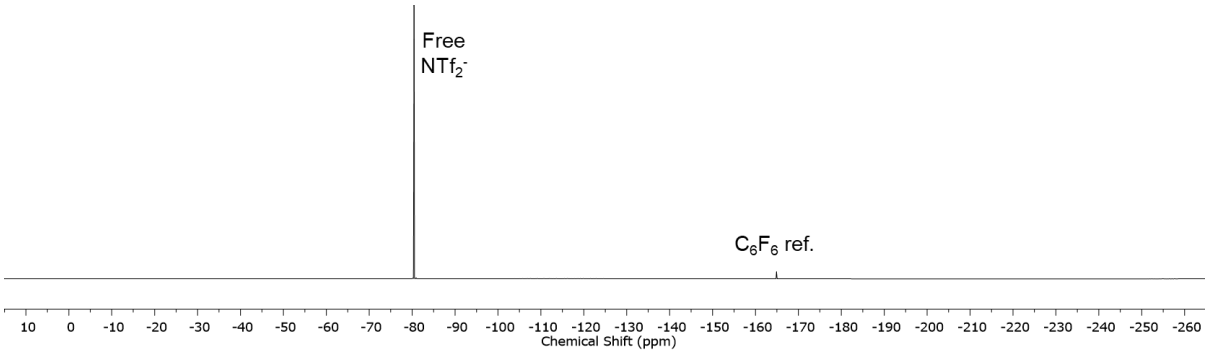
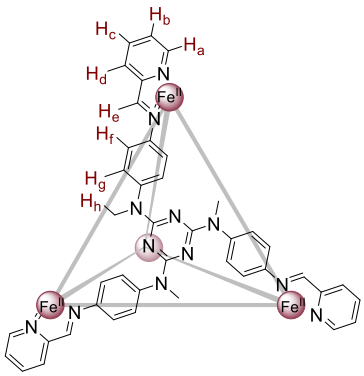


Figure 3.25 | ^{19}F NMR spectrum (471 MHz, 298 K, CD_3CN) of **3.1**.

The data was consistent with previously reported data for cage **3.1** with OTf^- counterions.[17]

Cage 3.2



N^2,*N*^{4},*N*^{6}-Tris(4-aminophenyl)-*N*^{2},*N*^{4},*N*^{6}-trimethyl-1,3,5-triazine-2,4,6-triamine **3.B** (50.0 mg, 0.1133 mmol, 1.0 equiv.) and 2-picolinaldehyde (32.3 µL, 0.3399 mmol, 3.0 equiv.) were dissolved in CH_3CN (10 mL) in a sealed 25 mL round bottom flask. The solution was degassed with N_2 for 10 min after which $Fe(NTf_2)_2$ (78.0 mg, 0.1133 mmol, 1.0 equiv.) was added. The solution was degassed for an additional 10 min. The solution was stirred at room temperature for 18 h. The dark purple solution was then concentrated under a flow of nitrogen. Addition of Et_2O (80 mL) precipitated the compound as a dark blue solid. The solid was separated by centrifugation and washed with Et_2O (2 × 80 mL). The solid was then dried under a flow of nitrogen.

^{1}H NMR (500 MHz, 298K, CD_3CN) – **3.2**: δ 8.84 (s, 12H, H_e), 8.50 (d, *J* = 7.7 Hz, 12H, H_d), 8.39 (t, *J* = 7.7 Hz, 12H, H_c), 7.75 (t, *J* = 7.7 Hz, 12H, H_b), 7.43 – 7.35 (bm, 36H, $H_{a,g}$), 5.24 – 4.89 (bm, 24H, H_f), 3.41 (s, 36H, H_h). **^{1}H NMR** (500 MHz, 298K, CD_3CN) – NTf_2^-⊂**3.2**: δ 8.80 (s, 12H, H_e), 8.47 (d, *J* = 7.7 Hz, 12H, H_d) 8.39 (t, *J* = 7.7 Hz, 12H, H_c), 7.75 (t, *J* = 7.7 Hz, 12H, H_b), 7.43 – 7.35 (bm, 36H, $H_{a,g}$), 5.24 – 4.89 (bm, 24H, H_f), 3.41 (s, 36H, H_h). **^{19}F NMR** (471 MHz, 298K, CD_3CN) δ –75.40 (NTf_2^-⊂**3.2**), –80.47 (free NTf_2^-). **LR-ESI-MS** [charge, calculated mass]: *m/z* = 1486.4 [**3.2**$(NTf_2)_5^{3+}$, 1486.5], 1044.2 [**3.2**$(NTf_2)_4^{4+}$, 1044.8], 779.7 [**3.2**$(NTf_2)_3^{5+}$, 779.8], 603.1 [**3.2**$(NTf_2)_2^{6+}$, 603.2], 476.9 [**3.2**$(NTf_2)^{7+}$, 477.0], 382.2 [**3.2**$^{8+}$, 382.3].

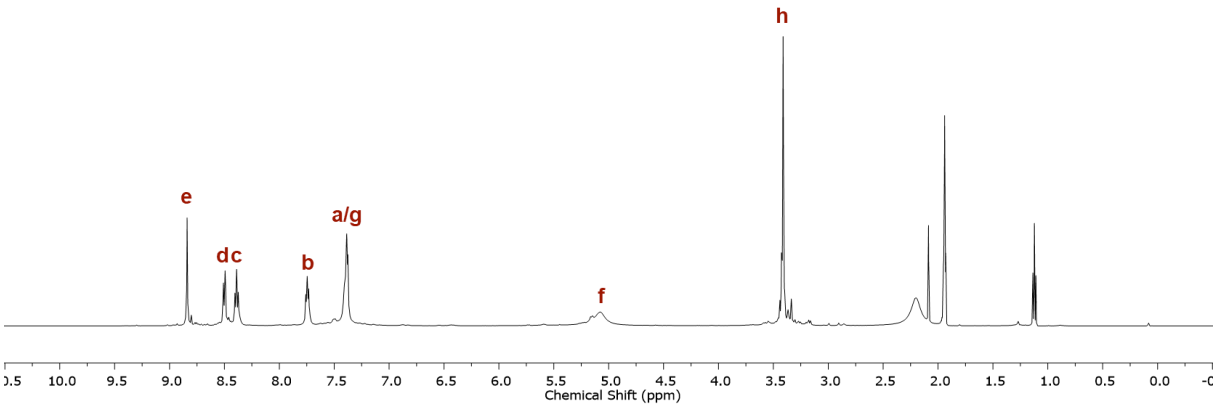


Figure 3.26 | ^{1}H NMR spectrum (500 MHz, 298 K, CD_3CN) of **3.2**.

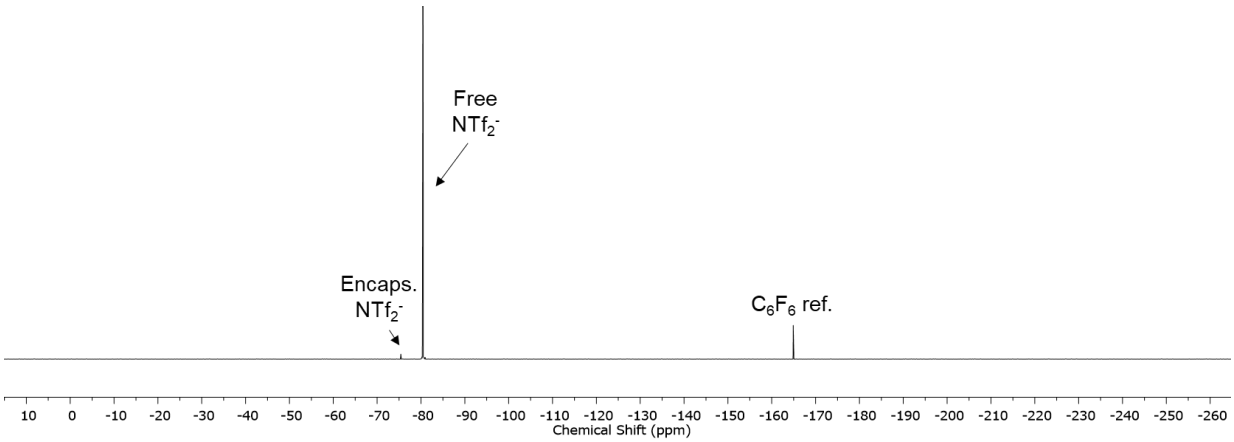


Figure 3.27 | ^{19}F NMR spectrum (471 MHz, 298 K, CD_3CN) of **3.2**.

The data was consistent with previously reported data for cage **3.2** with OTf^- counterions.[15]

3.6.2 General preparation for hybrid gels 3.1⊂Gel and 3.2⊂Gel

For the gels containing 1 mM of cage, a stock solution of 5 mM cage was prepared in CH_3CN. The peptide was dissolved in CH_3CN and added to the required volume of the cage solution to get the desired final concentrations of peptide (50 mM) and cage of (1 mM). The sample was sonicated for 5 min to promote the formation of the gel. For the gels containing 5 mM of cage, a solution of cage (5 mM) was prepared in CH_3CN. The peptide was directly added to the cage solution to get the desired final concentrations of peptide (50 mM) and cage (5 mM). The sample was sonicated for 10 min to promote formation of the gel.

All samples used in NMR studies were prepared and sealed in J-Young NMR tubes to prevent evaporation of the solvent over time.

3.6.3 Stability of 3.1 and 3.2

In solution

KTFA (35 µmol, 20 eq.) dissolved in CD_3CN (20 µL) was added to a solution of cage **3.1** or FA⊂**3.2** in CD_3CN (5 mM, 0.35 mL).

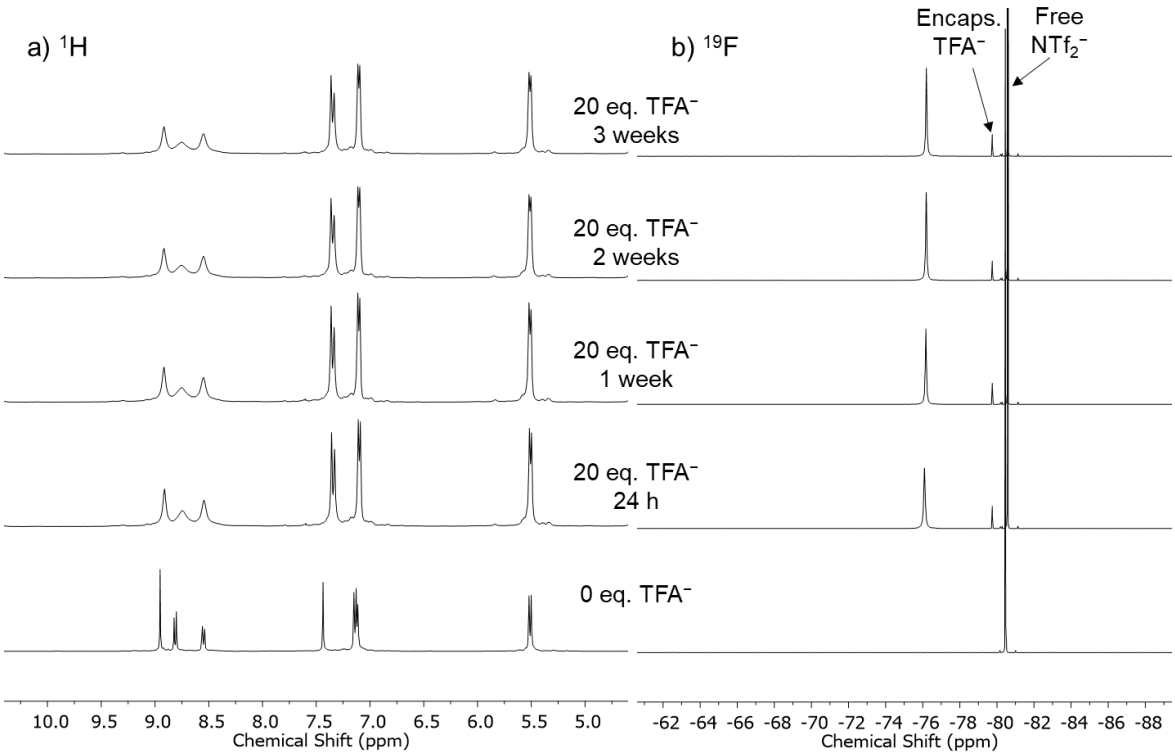


Figure 3.28 | a) ^1H NMR spectra (400 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (376 MHz, 298K, CD_3CN) of **3.1** in solution before and after addition of KTFA (20 eq.). Time elapsed from addition of KTFA is given on the spectra.

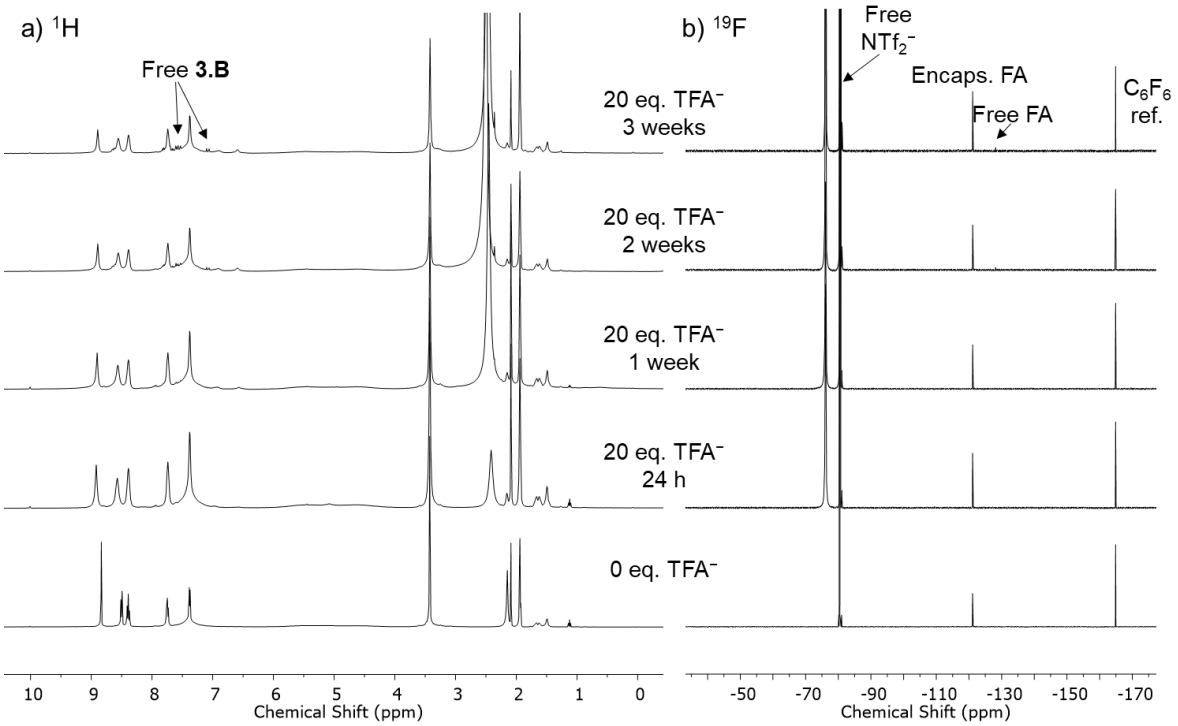


Figure 3.29 | a) ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (376 MHz, 298K, CD_3CN) of FA⊂**3.2** in solution before and after addition of KTFA (20 eq.). Time elapsed from addition of KTFA is given on the spectra.

In hybrid gels

3.C (12.5 mg, 17.5 µmol, 10 eq.) was added to a solution of cage **3.1** or FA⊂**3.2** in CD_3CN (5 mM, 0.35 mL). The samples were sonicated for 10 min to promote the formation of the gel.

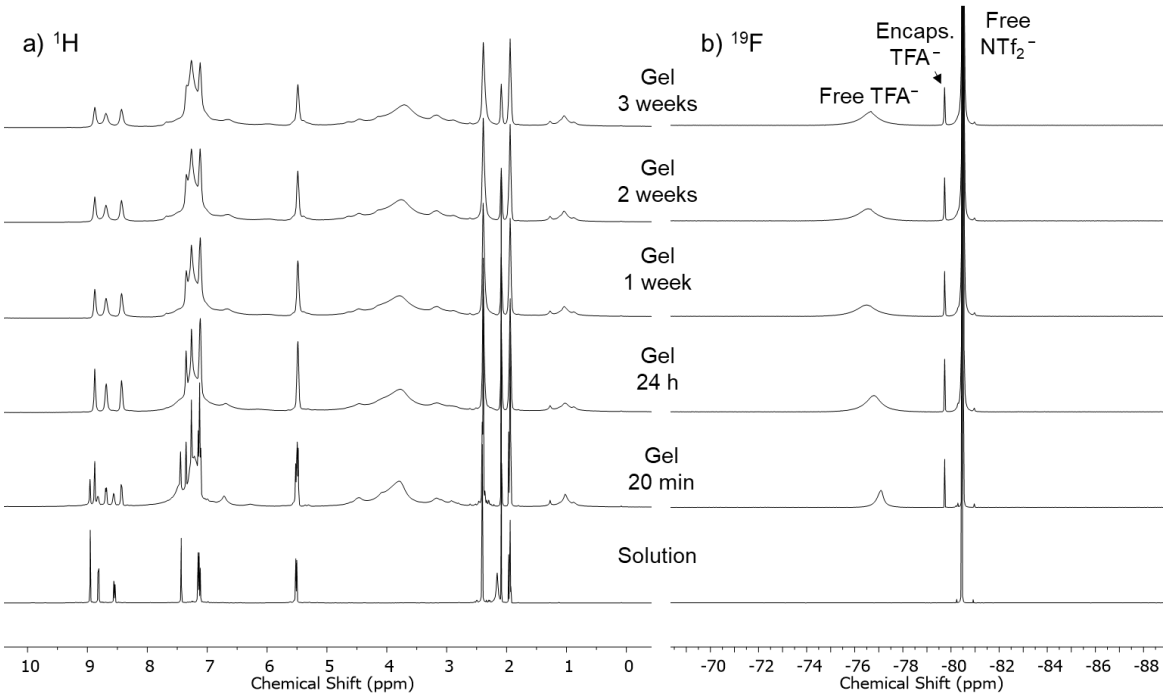


Figure 3.30 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (471 MHz, 298K, CD_3CN) of **3.1** before addition of **3.C** and gelation and of **3.1⊂Gel**. Time elapsed from gelation is given on the spectra.

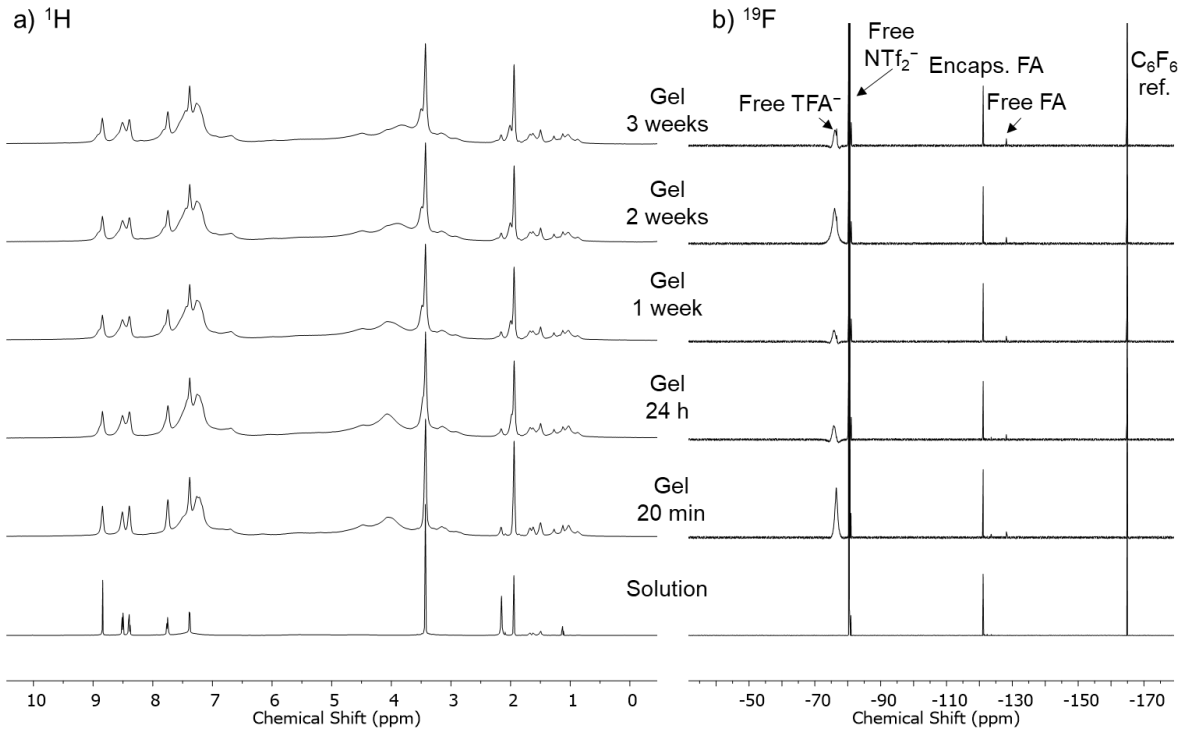


Figure 3.31 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (471 MHz, 298K, CD_3CN) of FA⊂**3.2** before addition of **3.C** and gelation and of FA⊂**3.2**⊂**Gel**. Time elapsed from gelation is given on the spectra.

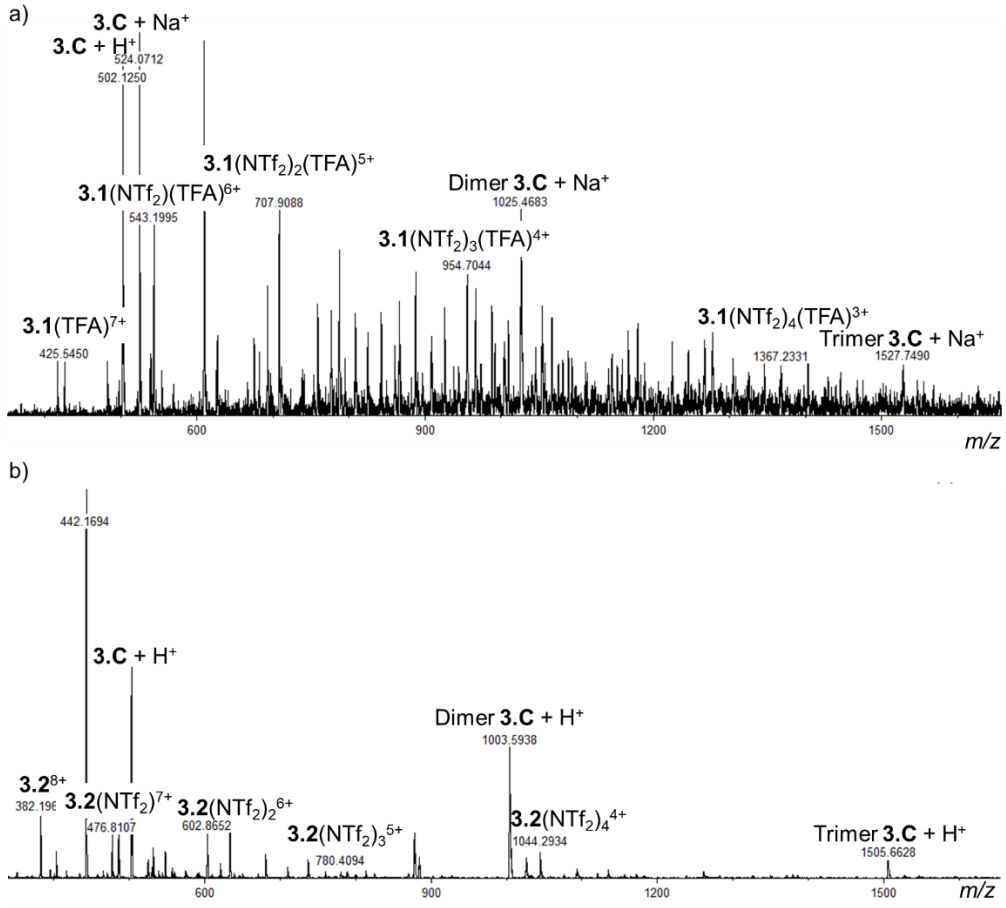


Figure 3.32 | LR-ESI-MS of a) **3.1** and b) **3.2** recovered from **3.1**⊂**Gel** and **3.2**⊂**Gel** respectively.

3.6.4 Guest uptake in 3.1 and 3.2

In solution

KTFA (35 µmol, 20 eq.) dissolved in CD_3CN (20 µL) was added to a solution of cage **3.1** or **3.2** in CD_3CN (5 mM, 0.35 mL). After 2 h of equilibration, $TBAReO_4$ or FA (1.75 µmol, 1 eq.) CD_3CN (20 µL) was added to solutions of **3.1** or **3.2** respectively.

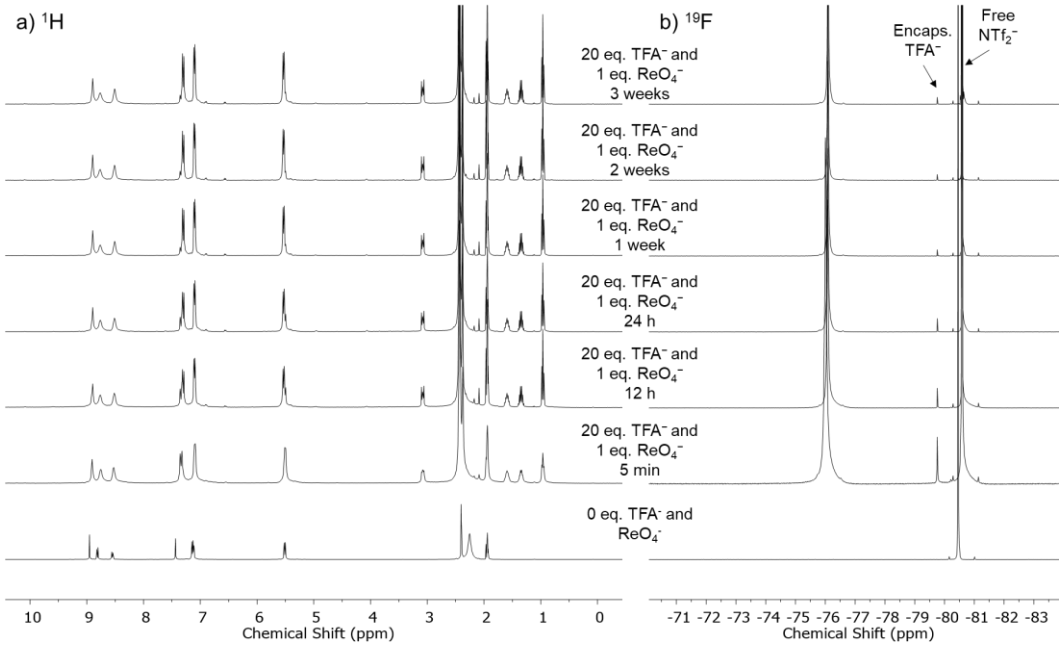


Figure 3.33 | a) ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (376 MHz, 298K, CD_3CN) of **3.1** in solution after addition of KTFA (20 eq.) followed by $TBAReO_4$ (1 eq.). Time elapsed from guest addition is given on the spectra.

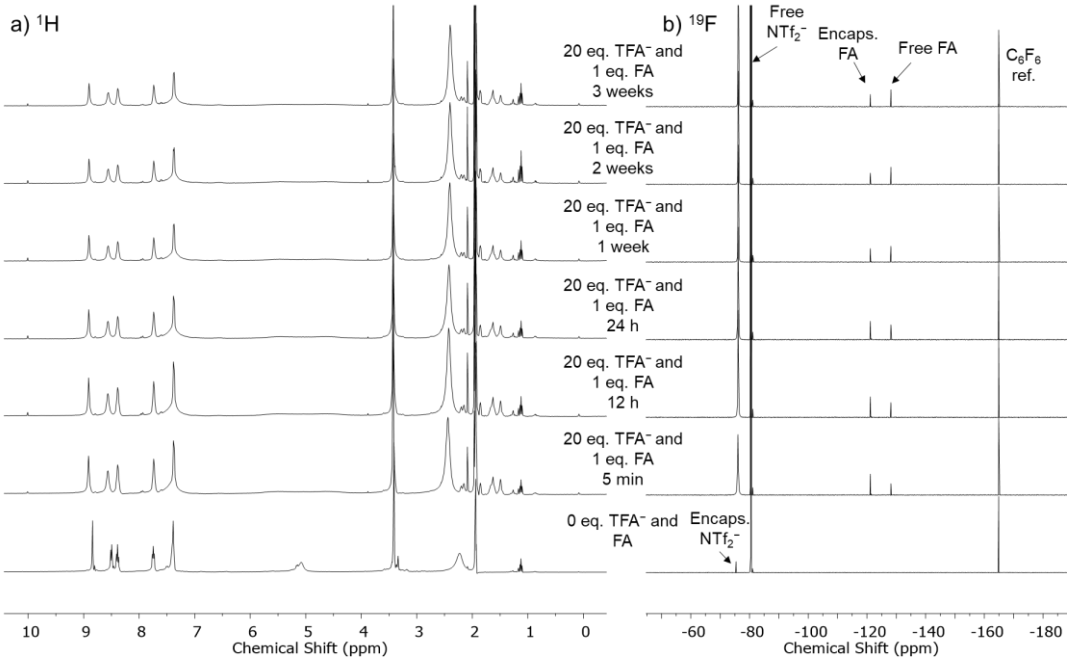


Figure 3.34 | a) ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (376 MHz, 298K, CD_3CN) of **3.2** in solution after addition of KTFA (20 eq.) followed by FA (1 eq.). Time elapsed from guest addition is given on the spectra.

In hybrid gels

3.C (12.5 mg, 17.5 µmol, 10 eq.) was added to a solution of cage **3.1** or **3.2** in CD_3CN (5 mM, 0.35 mL). The samples were sonicated for 10 min to promote the formation of the gel. After 2 h of equilibration, $TBAReO_4$ or FA (1.75 µmol, 1 eq.) CD_3CN (20 µL) was layered on top of **3.1⊂Gel** or **3.2⊂Gel** respectively.

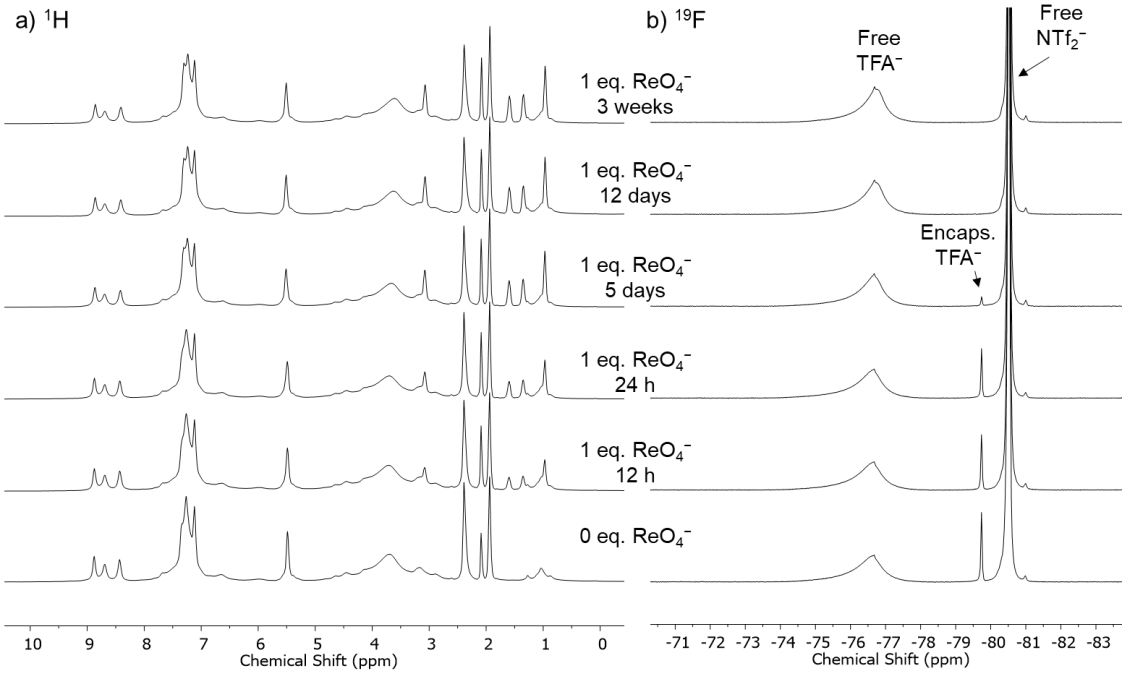


Figure 3.35 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (471 MHz, 298K, CD_3CN) of TFA^-⊂**3.1⊂Gel** before and after addition of $TBAReO_4$ (1 eq.). Time elapsed from guest addition is given on the spectra.

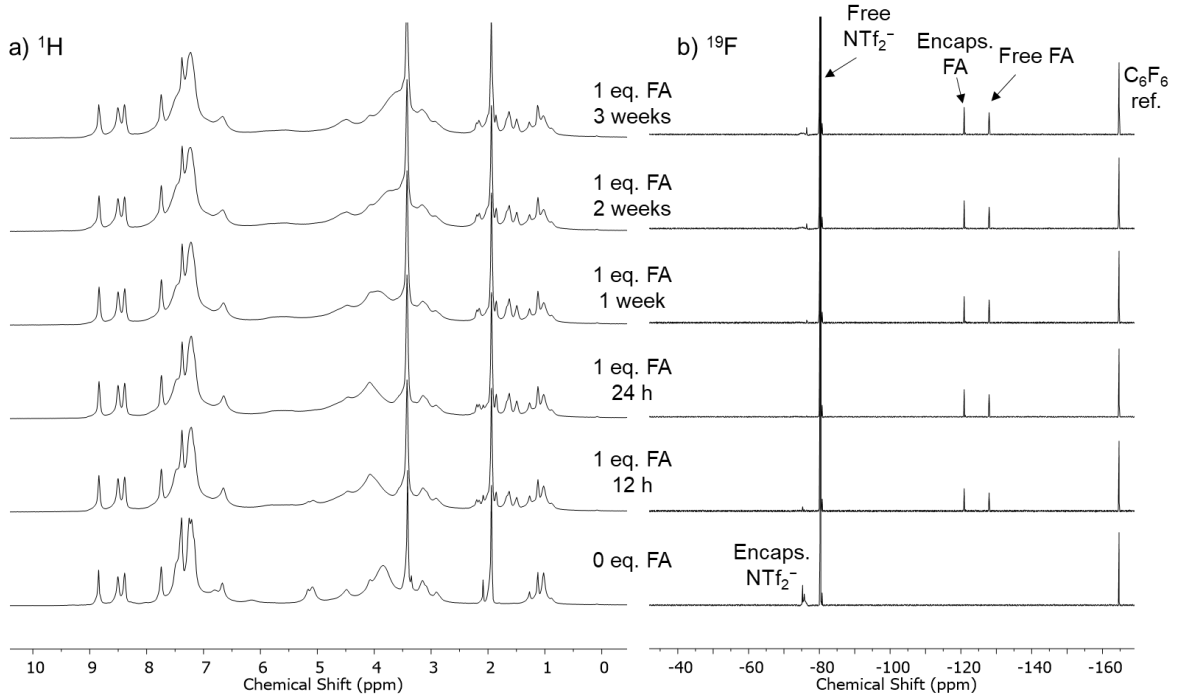


Figure 3.36 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (471 MHz, 298K, CD_3CN) of **3.2⊂Gel** before and after addition of FA (1 eq.). Time elapsed from guest addition is given on the spectra.

3.6.5 Diffusion of compounds in hybrid gels

FA in 3.1⊂Gel

FA (1 eq., 1.75 µmol) in CD_3CN (20 µL) was layered on top of ReO_4^-⊂**3.1**⊂**Gel**.

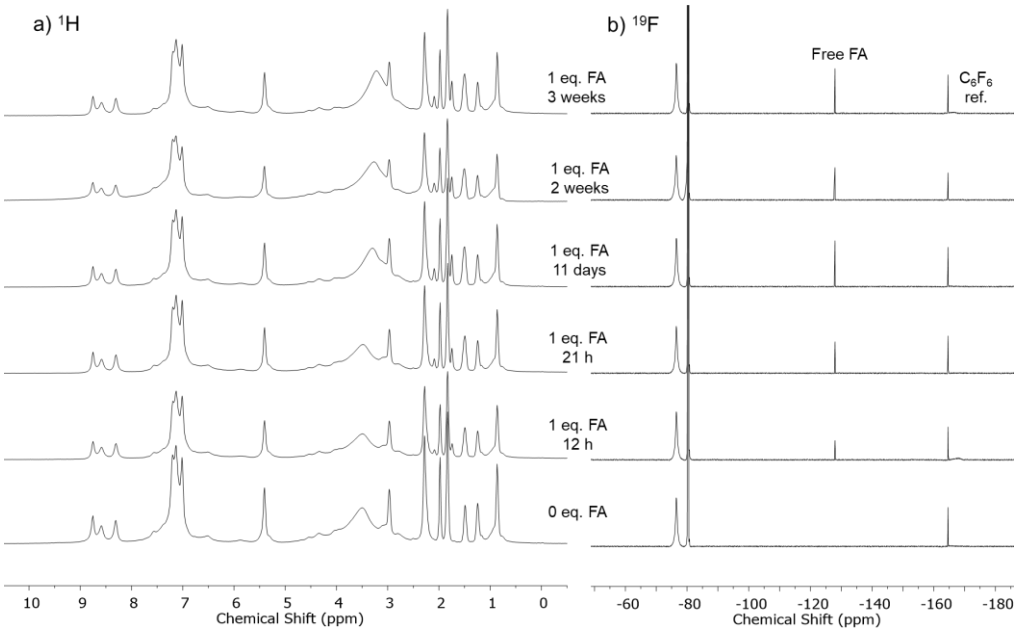


Figure 3.37 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (471 MHz, 298K, CD_3CN) of ReO_4^-⊂**3.1**⊂**Gel** before and after addition of FA (1eq.). Time elapsed from guest addition is given on the spectra.

Cages in 3.1⊂Gel and 3.2⊂Gel

A solution of TFA^-⊂**3.1** or FA⊂**3.2** (1 mM, 0.35 µmol) in CD_3CN (50 µL) was layered on top of FA⊂**3.2**⊂**Gel** or ReO_4^-⊂**3.1**⊂**Gel** respectively.

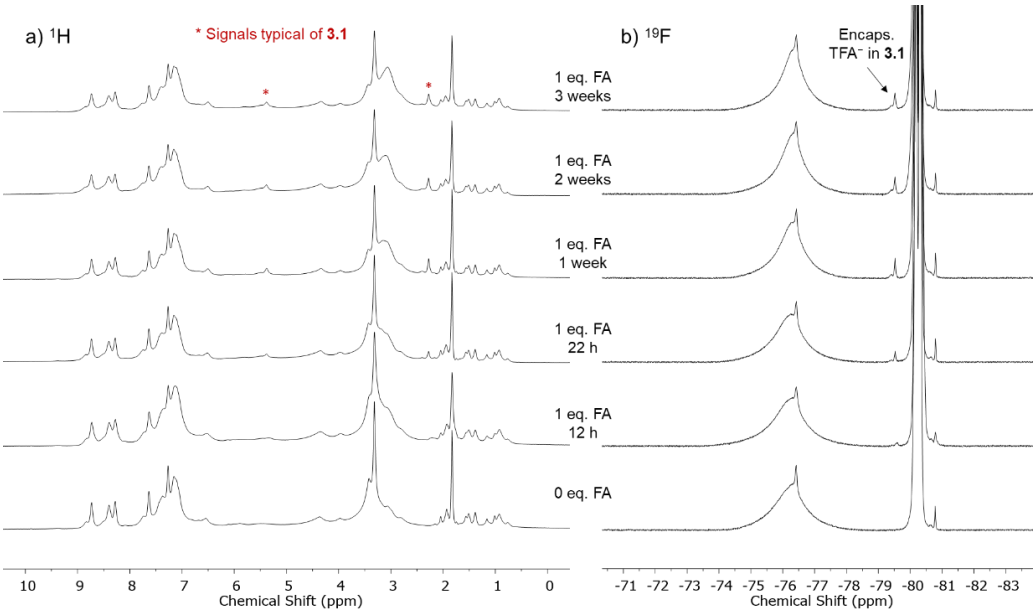


Figure 3.38 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (471 MHz, 298K, CD_3CN) of **3.2**⊂**Gel** after addition of TFA^-⊂**3.1**. Time elapsed from cage addition is given on the spectra.

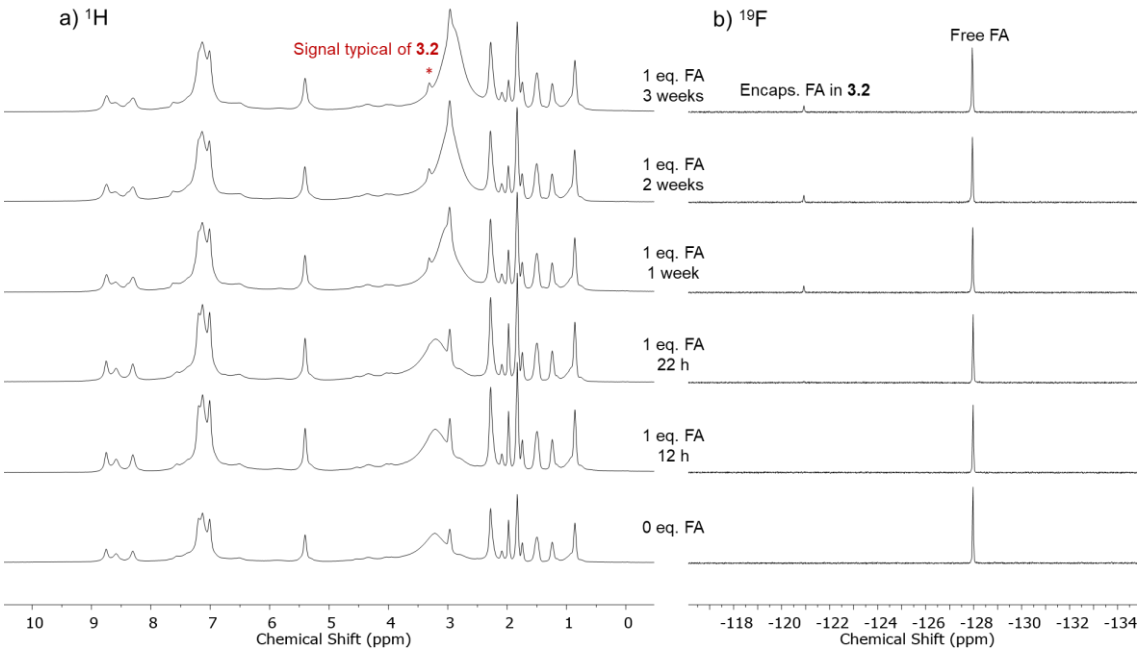


Figure 3.39 | a) ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) and b) ^{19}F NMR spectrum (471 MHz, 298K, CD_3CN) of **3.1⊂Gel** after addition of FA⊂**3.2**. Time elapsed from cage addition is given on the spectra.

3.6.6 Spatial segregation of cages and guest separation

The set up for slice selective NMR was similar to the one previously reported.[20]

Shaped pulse excitation was the Bruker standard gaussian-cascade G4 using:

NS = 8

D1 = 1.00 s

SPW1 = 0.30642

P11 = 1494

GPZ1 = 23.5%

^{1}H and ^{19}F NMR 2D mapping of the sample were used to localize the offset corresponding to each layer of gel under the conditions (slice width) used for the 1D experiments. Fewer and thicker slices were used (approx. 2 mm, 10 slices) than in the initial 2D mapping in order to get better intensity of the signals. The mapping was repeated with the following parameters:

^{1}H NMR:

GPZ1 = 6%, SPW1 = 0.30642, P11 = 1494, NS = 8

^{19}F NMR:

GPZ1 = 11.5%, SPW1 = 4.792, P11 = 795.2, NS = 8

The optimal slices showing most intense signal in the ^{1}H NMR were identified for each layer of gel and the SPOFFS1 were recorded:

Top layer: SPOFFS1 = 9200 Bottom layer: SPOFFS1 = –9200

The SPOFFS1 being dependent of the frequency of excitation and the probe used, the SPOFFS1 for the ^{19}F NMR were identified by matching the slices, giving:

Top layer: SPOFFS1 = 9200 Bottom layer: SPOFFS1 = –18400

In order to keep the performance of the ^{19}F NMR experiments similar to those of the ^{1}H NMR experiments, a 20 ppm spectral width was acquired and offset selected to observe each region of interest. One region of the ^{1}H NMR (0 – 10 ppm, NS = 8) and two regions of the ^{19}F NMR (−135 to −115 ppm and −88 to −68 ppm, NS = 256) were monitored. This meant that it was not possible to get quantitative data as peaks could not be integrated against a reference in the ^{19}F spectrum. However, relative values can be obtained by comparing the absolute values of the integrals of peaks between different time points as the gain and number of scans were kept identical between runs.

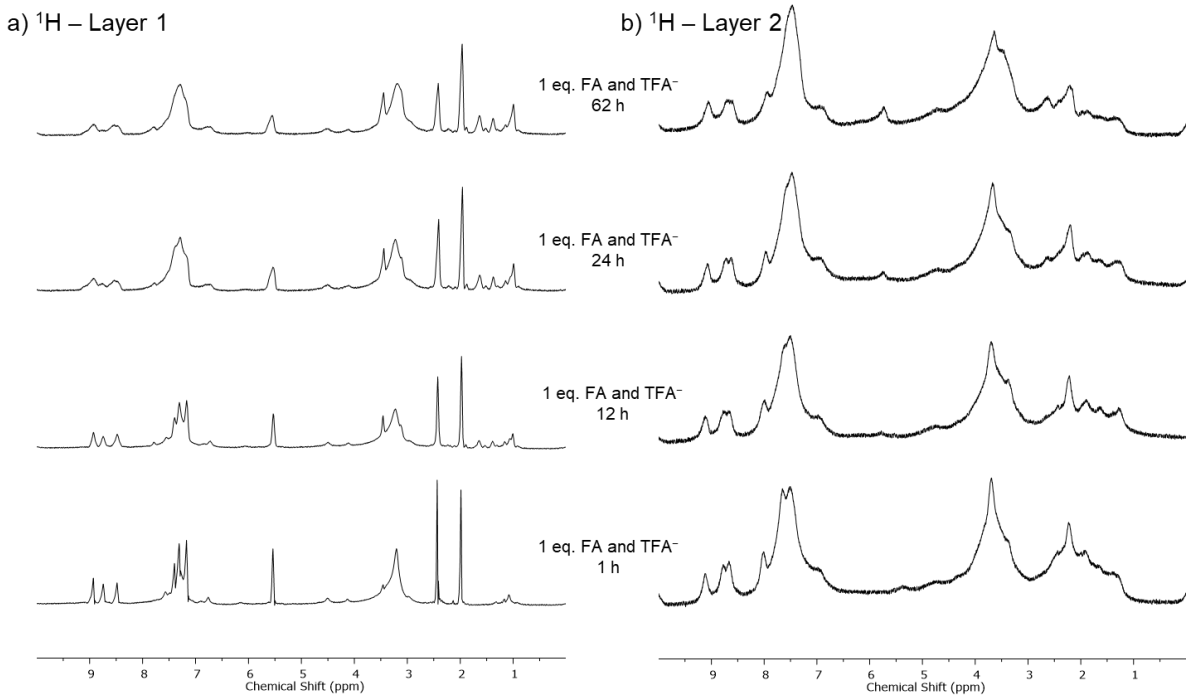


Figure 3.40 | ^{1}H NMR spectra (500 MHz, 298 K, CD_3CN) of a) the bottom layer containing **3.1** and b) the top layer containing **3.2** over 62 h after addition of a mixture of FA and $TBAReO_4$ (1 eq.).

3.7 References

[1] C. D. Jones, J. W. Steed, *Chem. Soc. Rev.* **2016**, *45*, 6546-6596.

[2] B. O. Okesola, D. K. Smith, *Chem. Soc. Rev.* **2016**, *45*, 4226-4251.

[3] J. Mayr, C. Saldías, D. Díaz Díaz, *Chem. Soc. Rev.* **2018**, *47*, 1484-1515.

[4] N. Hosono, S. Kitagawa, *Acc. Chem. Res.* **2018**, *51*, 2437-2446.

[5] W. Zheng, G. Yang, N. Shao, L.-J. Chen, B. Ou, S.-T. Jiang, G. Chen, H.-B. Yang, *J. Am. Chem. Soc.* **2017**, *139*, 13811-13820.

[6] M. Mauro, *Eur. J. Inorg. Chem.* **2018**, *2018*, 2090-2100.

[7] A. Carné-Sánchez, G. A. Craig, P. Larpent, T. Hirose, M. Higuchi, S. Kitagawa, K. Matsuda, K. Urayama, S. Furukawa, *Nat. Commun.* **2018**, *9*, 2506.

[8] T.-H. Chen, L. Wang, J. V. Trueblood, V. H. Grassian, S. M. Cohen, *J. Am. Chem. Soc.* **2016**, *138*, 9646-9654.

[9] S. Ganta, D. K. Chand, *Inorg. Chem.* **2018**, *57*, 3634-3645.

[10] J. A. Foster, R. M. Parker, A. M. Belenguer, N. Kishi, S. Sutton, C. Abell, J. R. Nitschke, *J. Am. Chem. Soc.* **2015**, *137*, 9722-9729.

[11] M. Amit, S. Yuran, E. Gazit, M. Reches, N. Ashkenasy, *Adv. Mater.* **2018**, *30*, 1707083.

[12] A. Lampel, R. V. Ulijn, T. Tuttle, *Chem. Soc. Rev.* **2018**, *47*, 3737-3758.

[13] S. Marchesan, L. Waddington, C. D. Easton, D. A. Winkler, L. Goodall, J. Forsythe, P. G. Hartley, *Nanoscale* **2012**, *4*, 6752-6760.

[14] A. M. Garcia, D. Iglesias, E. Parisi, K. E. Styan, L. J. Waddington, C. Deganutti, R. De Zorzi, M. Grassi, M. Melchionna, A. V. Vargiu, S. Marchesan, *Chem* **2018**, *4*, 1862-1876.

[15] J. L. Bolliger, T. K. Ronson, M. Ogawa, J. R. Nitschke, *J. Am. Chem. Soc.* **2014**, *136*, 14545-14553.

[16] A. M. Castilla, N. Ousaka, R. A. Bilbeisi, E. Valeri, T. K. Ronson, J. R. Nitschke, *J. Am. Chem. Soc.* **2013**, *135*, 17999-18006.

[17] A. M. Castilla, M. A. Miller, J. R. Nitschke, M. M. J. Smulders, *Angew. Chem. Int. Ed.* **2016**, *55*, 10616-10620.

[18] A. B. Grommet, J. L. Bolliger, C. Browne, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2015**, *54*, 15100-15104.

[19] N. Busschaert, S. J. Bradberry, M. Wenzel, C. J. E. Haynes, J. R. Hiscock, I. L. Kirby, L. E. Karagiannidis, S. J. Moore, N. J. Wells, J. Herniman, G. J. Langley, P. N. Horton,

M. E. Light, I. Marques, P. J. Costa, V. Félix, J. G. Frey, P. A. Gale, *Chem. Sci.* **2013**, *4*, 3036-3045.

[20] A. B. Grommet, J. B. Hoffman, E. G. Percástegui, J. Mosquera, D. J. Howe, J. L. Bolliger, J. R. Nitschke, *J. Am. Chem. Soc.* **2018**, *140*, 14770-14776.

[21] A.-C. Pöppler, S. Frischkorn, D. Stalke, M. John, *ChemPhysChem* **2013**, *14*, 3103-3107.

[22] E. G. Percástegui, J. Mosquera, T. K. Ronson, A. J. Plajer, M. Kieffer, J. R. Nitschke, *Chem. Sci.* **2019**, *10*, 2006-2018.

Chapter 4

A tripeptide functionalised $Fe^{II}_4L_4$ tetrahedral cage undergoes light induced reversible gelation

4.1 Introduction

Assembled in living organisms or by chemists to generate peptides or proteins, amino acids are the building blocks of life.[1] Differentiated by their specific side chains, they have a general chemical structure containing both a carboxylic acid (C-terminus) and an amine (N-terminus) functional group. Based on the amino-acid sequence employed, peptides and proteins are able to self-organise into higher order structures and materials.[2-5] Peptides are truly versatile building blocks, and new properties could emerge from their use in conjunction with supramolecular coordination structures.[6-8]

The bioconjugation of peptides onto coordination cages thus offers a straightforward strategy to enhance the properties of the cages. This strategy synergistically combines the characteristics of both cages (defined 3D-structure and cavity, host-guest chemistry…) and the peptides used as building blocks. For example, we can envision that once grafted with a peptide gelator, a cage will form a network of fibrils leading to the gelation of the bulk material, while preserving the cage structure.

Previous results by Li *et al.* showed that sol-gel transition of a Phe-Phe dipeptide could be controlled by light in the presence of a photo-acid generator (PAG).[9] Upon exposure to visible light, merocyanine (the protonated form of the PAG), undergoes a ring-closing reaction to form the spiropyran, releasing a proton into the environment. The decrease of the pH in bulk solution disturbs the packing of the peptide fibrils, leading to the return of the sample to its liquid state. Based on this work, we hypothesised that the sol-gel transition of a gel formed from cages grafted with peptide gelators could be triggered by an external stimulus in a reversible process.

This chapter explores the formation of coordination cages which incorporate amino acids or short peptides as building blocks. While direct conjugation through the N-terminus was unsuccessful, both amino-acids and peptides were appended onto a $Fe^{II}_4L_4$ tetrahedral cage *via* an aniline linker. Building on results from chapter 3, a gel was formed from the cage grafted with the peptide gelator **3.C**. Combining this cage with a photoacid generator was predicted to lead to the disassembly of the peptide fibrils, while preserving the integrity of the coordination cage. The reversible gelation of peptide-cage system could thus be achieved by exposing or isolating the sample from light.

4.2 Functionalisation of cages with amino acids

The N-terminus of amino acids and peptides could potentially react with picolinaldehyde to form pyridyl-imines, leading to the formation of supramolecular architectures after the addition of metal salts. The direct self-assembly of **3.A** with various hydrophobic amino acids (glycine, alanine, valine, leucine, isoleucine and phenylalanine) and $Fe(NTf_2)_2$ was attempted in CH_3CN (Figure 4.2). Methylated amino acids were also employed to increase the solubility of the compounds formed in organic solvents. No discrete supramolecular species could be detected by ^1H NMR or ESI-MS. Imine condensation occurred between **3.A** and the methylated amino acids in CH_2Cl_2, compounds which could be isolated and dissolved in CH_3CN. Unfortunately, no discrete supramolecular structures were observed by ^1H NMR after the addition of $Fe(NTf_2)_2$ (Figure 4.2 d). In both attempted self-assemblies (direct or with preformation of the imine), only the signals for the amino acids were observed in the ^1H NMR. Other metals (Zn^{II} and Co^{II}) as well as anions (OTf^-, BF_4^-, ClO_4^-) were screened for self-assembly without success.

The MM3-optimised model of the expected tetrahedral capsule showed the close proximity of the C-termini of the amino acids (Figure 4.1). We hypothesised that the steric clash between the carboxylic acids of the amino acid around the metal vertices was too high and therefore prevented self-assembly into a defined 3D architecture.

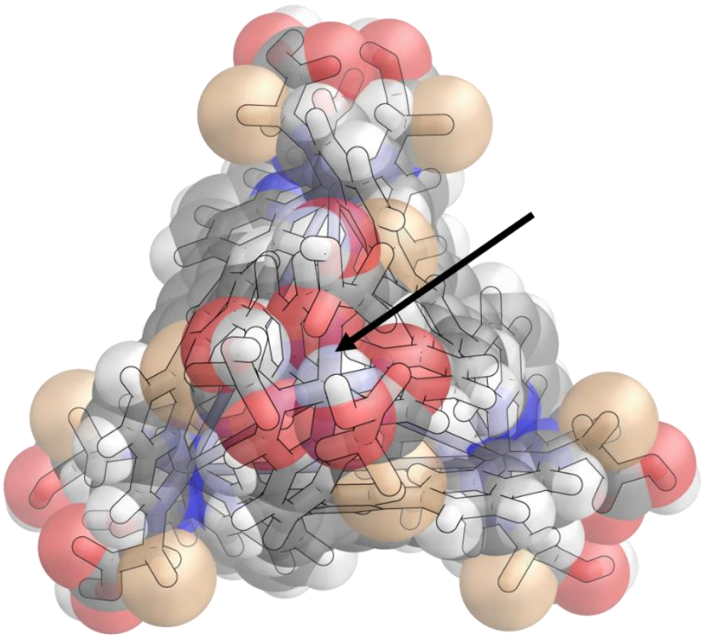


Figure 4.1 | MM3-optimised model of the tetrahedral capsule expected from amino-acids (position residues in gold) and **3.A**. The space filling highlights the steric clash between carboxylic acids around the metal vertices (black arrow).

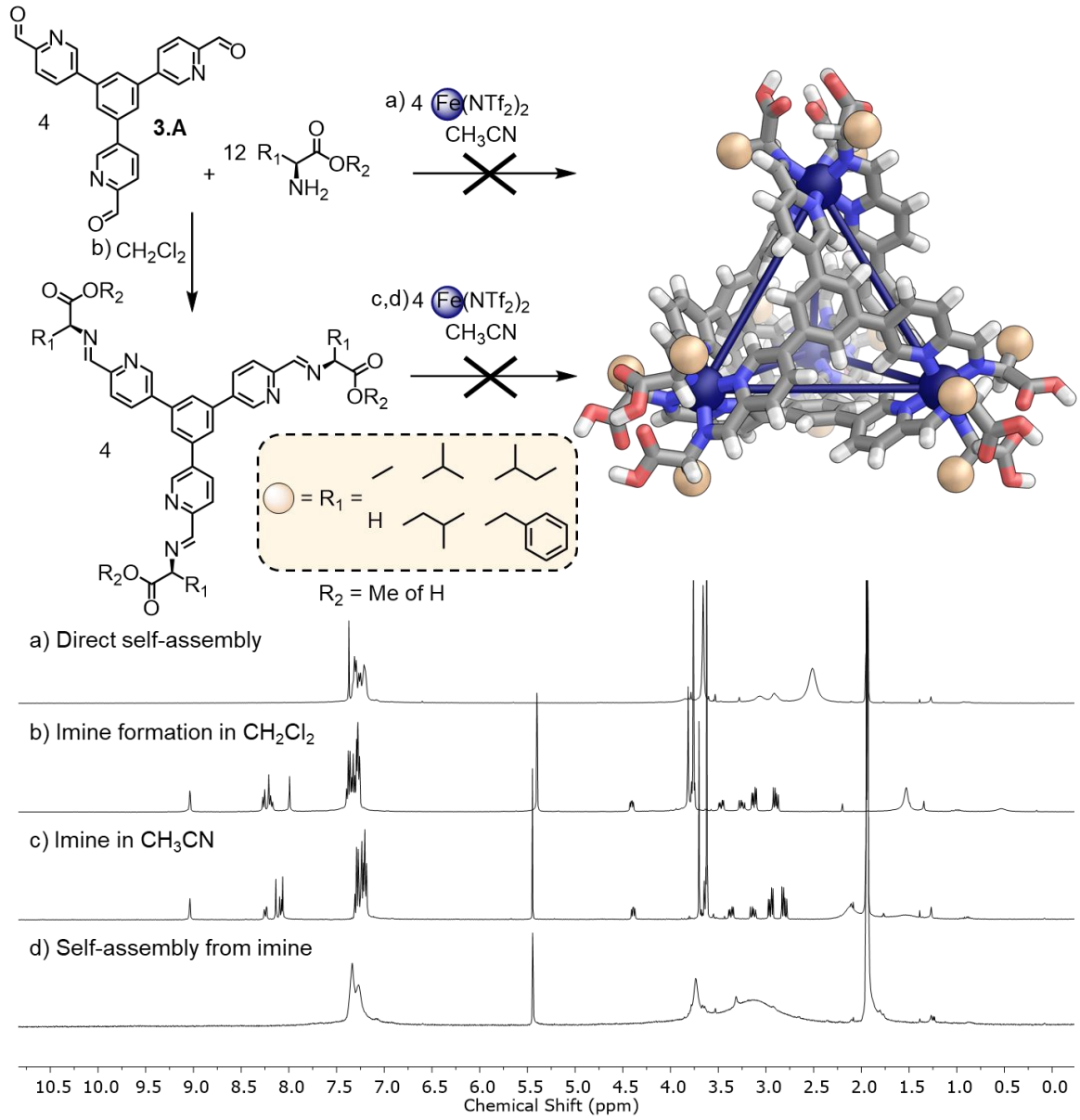
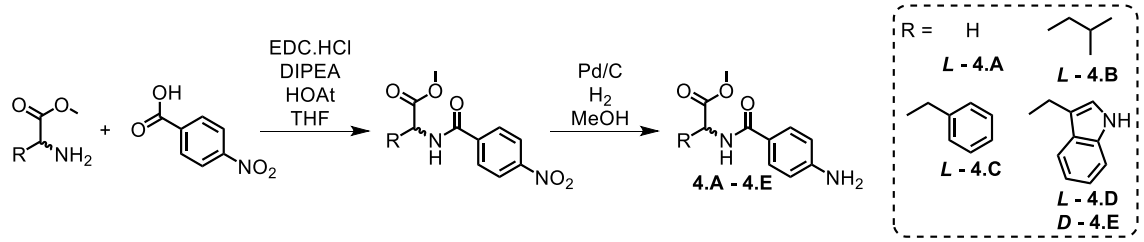


Figure 4.2 | Top: Synthetic scheme for the formation of a tetrahedral capsule from subcomponent **3.A** and amino acids (glycine, alanine, valine, leucine, isoleucine and phenylalanine) through a) direct self-assembly or b) via preformation of the tri-imine followed by c) transfer into CH_3CN and d) addition of $Fe(NTf_2)_2$. MM3-optimised model of the expected tetrahedral capsule. Bottom: ^1H NMR spectra (400 MHz, 298 K, CD_3CN (a, c, d) and CH_2Cl_2 (b)) of a) the attempted direct self-assembly of **3.A** and phenylalanine methyl ester with $Fe(NTf_2)_2$, b) the tri-imine formed from **3.A** and phenylalanine methyl ester, c) the tri-imine in CD_3CN and d) the attempted self-assembly from the preformed imine after addition of $Fe(NTf_2)_2$.

To increase the distance between amino acid residues at the vertices, compounds **4.A** – **4.E** incorporating an aniline were synthesised following Scheme 4.1. The reaction of anilines **4.A** – **4.E** (12 eq.) with **3.A** (4 eq.) and $Fe(NTf_2)_2$ (4 eq.) in CH_3CN led to the formation of $Fe^{II}_4L_4$ assemblies **4.1** – **4.5** respectively, after heating for 18 h at 50 °C under N_2, as confirmed by ESI-MS. The ^1H NMR spectra indicated the presence of two sets of signals, each corresponding to a highly symmetrical species in solution (Figure 4.3).



Scheme 4.1 | General synthetic scheme for the synthesis of compounds **4.A – 4.E.**

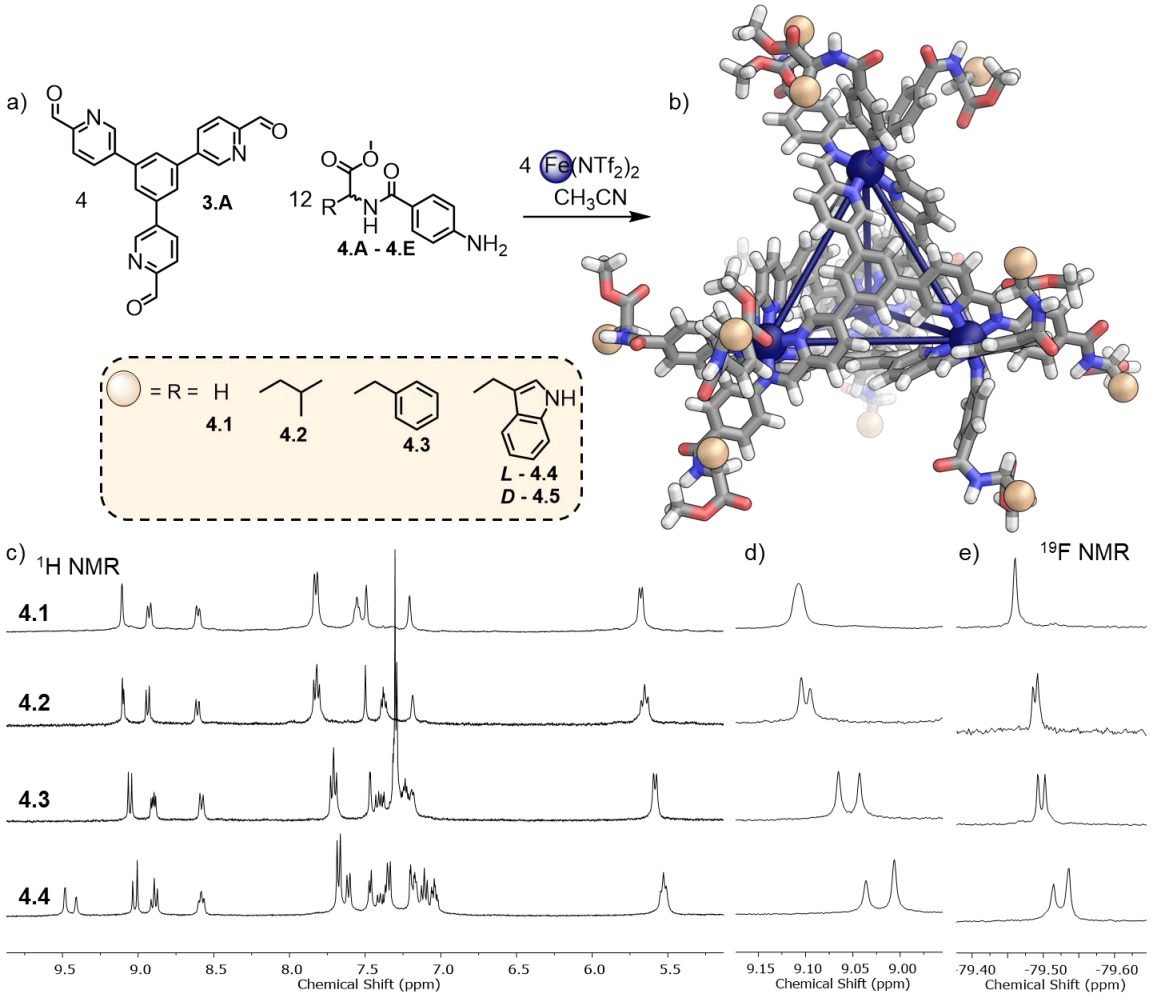


Figure 4.3 | a) Synthetic scheme for the formation of tetrahedral capsules **4.1 – 4.5** from subcomponents **3.A**, anilines **4.A – 4.E** (glycine, leucine, phenylalanine and tryptophan) and $Fe(NTf_2)_2$ in CH_3CN. b) MM3-optimised model of tetrahedral capsule formed. c) Aromatic region of the ^1H NMR spectra (400 MHz, 298 K, CD_3CN) for **4.1 – 4.5**, d) zoom on the imine region of the ^1H NMR spectra and e) ^{19}F NMR spectra (376 MHz, 298 K, CD_3CN) after addition of KTFA (1 eq.) zoomed in the region corresponding to $TFA^-\subset$**4.1** – **4.5**.

T-symmetric tetrahedra are the only observed symmetry for capsules formed from **3.A**, for which the ΛΛΛΛ and the ΔΔΔΔ conformation of metal vertices are both present in solution as an enantiomeric pair. The splitting observed in the NMR was therefore attributed to the formation of diastereisomers (Δ-L and Λ-L) when the chiral aniline was incorporated in tetrahedral capsules. Interestingly, depending on the amino acid present on the aniline, the ratio of diastereoisomers observed was different (Figure 4.3 c, d) which was inferred to result from

the chiral aniline influencing the ratio of ΛΛΛΛ or ΔΔΔΔ isomer of the cage formed. Furthermore, upon addition of a fluorinated molecule which can act as a guest for **4.1 – 4.5**, here TFA^-, the signal for the guest was also observed to split (Figure 4.3 e), indicating that the chirality of the cage is felt by the guest within the cavity.

The degree of chiral induction in **4.1 – 4.5** was further probed by circular dichroism (CD). The CD response was normalised by the concentration of the cages determined by UV-vis spectroscopy so that the response was directly proportional to the enantiomeric excess of the compound. Cage **4.1** made from the achiral aniline (glycine methyl ester) showed no CD signal whereas **4.2 – 4.5** showed signals of varying intensities (Figure 4.4 a).The bands in the 500 – 650 nm are characteristic of the MLCT bands of the cages and therefore relate directly to the chirality of the metal vertices (Figure 4.4 b). The trend observed for the intensity of the CD signal was similar to that observed in the NMR, showing the greater chirality change induced by **4.D** (L-tryptophan methyl ester) relative to **4.C** (L-leucine methyl ester) and **4.B** (L-phenylalanine methyl ester). Furthermore, the CD response observed for **4.D** (L-tryptophan methyl ester) and **4.E** (D-tryptophan methyl ester) was of opposite sign (Figure 4.4 c). Thus, when using amino-acids of opposite chirality induced opposite chiral enrichment of the cages.

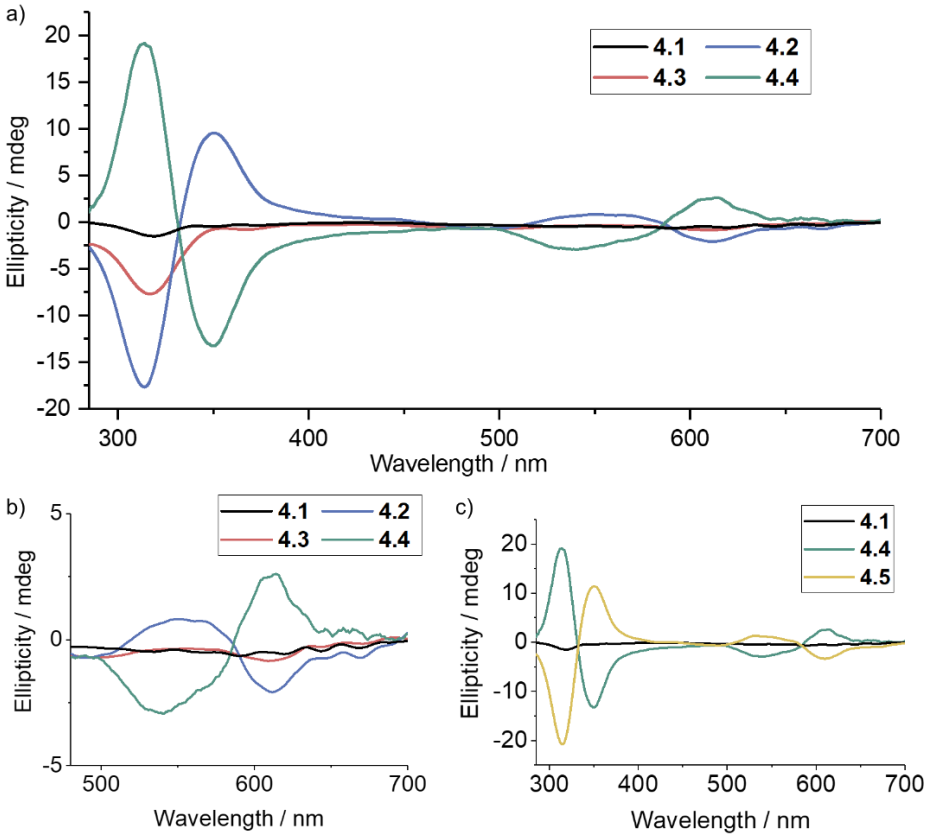


Figure 4.4| a) Circular dichroism (CD) spectra of cages **4.1 – 4.5**. b) Zoom in the MLCT band of the CD spectra for **4.1 – 4.5** showing the range of induction on the cages by different amino acids. c) CD spectra of cages **4.1**, **4.4** and **4.5** showing the opposite response of **4.4** and **4.5.**

4.3 Functionalisation of coordination cages with a peptide gelator

Building on the possibility of appending amino acids onto cages *via* an aniline, it was hypothesised that short peptides could also be conjugated to cages using the same approach.

Subcomponent **3.C** ((*p*-aminobenzoyl)-L-Phe-D-Ala-L-Phe-NH_2) introduced in Chapter Three was investigated for use as a subcomponent in self-assembly with **3.A**. The reaction of **3.C** (12 eq.) with **3.A** (4 eq.) and $Fe(NTf_2)_2$ (4 eq.) in CH_3CN led to the formation of $Fe^{II}_4L_4$ assemblies **4.6**, after heating for 18 h at 50 °C under N_2, as confirmed by NMR and ESI-MS (Figure 4.5 a and Figure 4.43). Similarly to **4.1 – 4.5**, the ^{1}H NMR spectrum of **4.6** indicated the presence of two sets of signals, assigned to the diastereoisomers of the compound (Figure 4.5 c). The difference in intensity of these signals was attributed to chiral induction exerted on the cage as confirmed by circular dichroism (Figure 4.5 d, e). Due to the presence of residual TFA^- with peptide **3.C**, the cage was always obtained as the host-guest complex with one TFA^- encapsulated in **4.6**. For ease of notation, $TFA^-\subset$**4.6** will be simplified to **4.6.**

Based on the ability of **3.C** to act as a low molecular weight gelator in CH_3CN, the gelling of **4.6** was tested at different concentrations. Cage **4.6** was obtained as a dark blue powder by precipitation of the cage from CH_3CN with Et_2O. The solid was dissolved at a known concentration in CH_3CN. Gentle heating was used to solubilise the cage if needed. Sonication for 1 h was necessary to promote the formation of a gel. Visible aggregates were observed from 2 mM. A final concentration above 5 mM was necessary to give a gel-like material which remained at the bottom of the sample tube (Figure 4.6).

Characterisation of the material by cryo-TEM is currently under investigation in collaboration with Giorgio Divitini (Materials Science Department, University of Cambridge). The response of the material to stress is under investigation by the Marchesan group in Trieste.

Preliminary TEM results indicated that the gel formed was sensitive to variations of its environment. We therefore hypothesised that the sol-gel transition could be triggered by an external stimulus leading to potential reversible process.

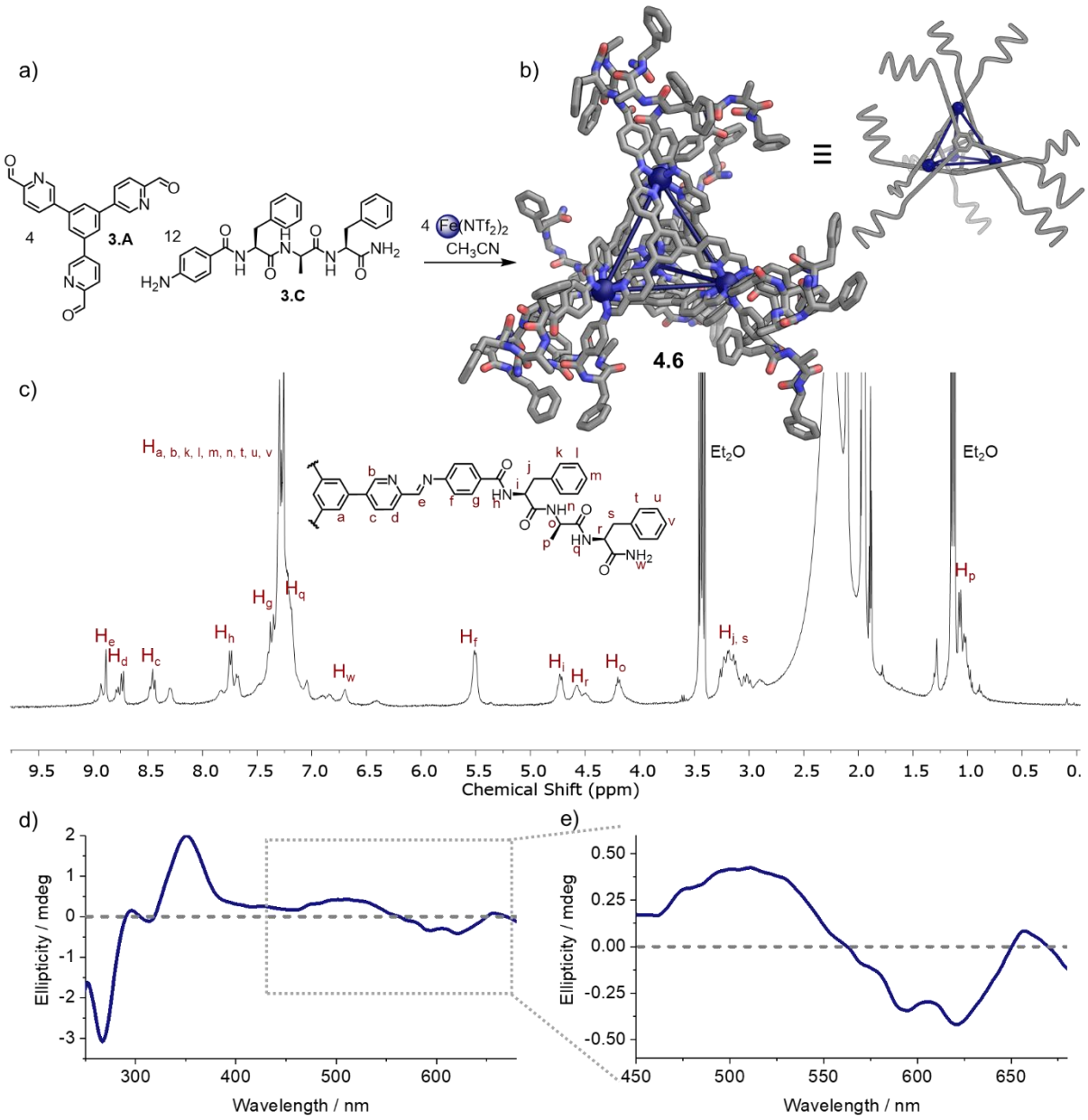
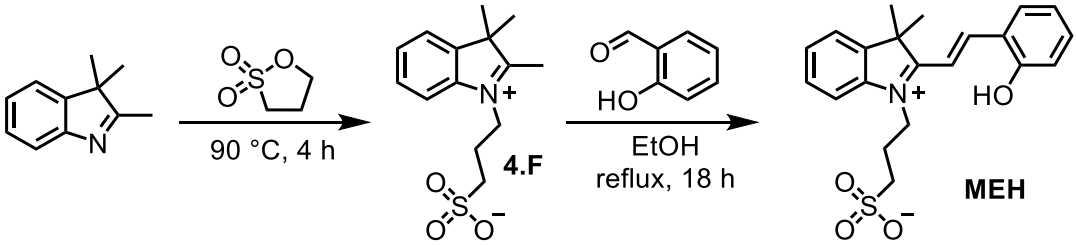


Figure 4.5 | a) Synthetic scheme for the formation of tetrahedral capsule **4.6** from subcomponents **3.A**, aniline **3.C** and $Fe(NTf_2)_2$ in CH_3CN. b) MM3-optimised model of tetrahedral capsule **4.6** and a schematic representation of the cage. c) ^1H NMR spectrum (400 MHz, 298 K, CD_3CN) of **4.6**. d) CD spectrum of cage **4.6**. e) Zoom in the MLCT band of the CD spectrum for **4.6** showing the induction on the cages by **3.C**.

Figure 4.6 | Gel formed from **4.6** at 5 mM in CH_3CN after sonication for 1 h.

4.4 Probing the compatibility of photoacid generator and cages

Based on the work by Li *et al.*, **MEH** was chosen as a PAG based on its metastable character which leads to long lived forms of both the **SP** (spiropyran) and **MEH** (merocyanine) form. (*E*)-3-(2-(2-Hydroxystyryl)-3,3-dimethyl-3*H*-indol-1-ium-1-yl)propane-1-sulfonate **MEH** was synthesised by Shihao Zang in two steps following a literature procedure (Scheme 4.2).[10]



Scheme 4.2 | Synthetic scheme for the synthesis of **MEH.**

Preliminary solution-phase experiments were conducted using cage **4.3** to verify the compatibility of the basic cage scaffold with the PAG. Both the UV-vis (Figure 4.7) and NMR spectra (Figure 4.8) of **4.3** and **4.6** were nearly identical, showing the similarity in the chemical properties of both cages. The structure of **4.3** is very similar to that of **4.6**, with the two cages only differing by the two additional amino acids at the periphery of the structure (L-Phe-OMe vs L-Phe-D-Ala-L-Phe). We inferred that this difference would not influence the properties of the cage framework.

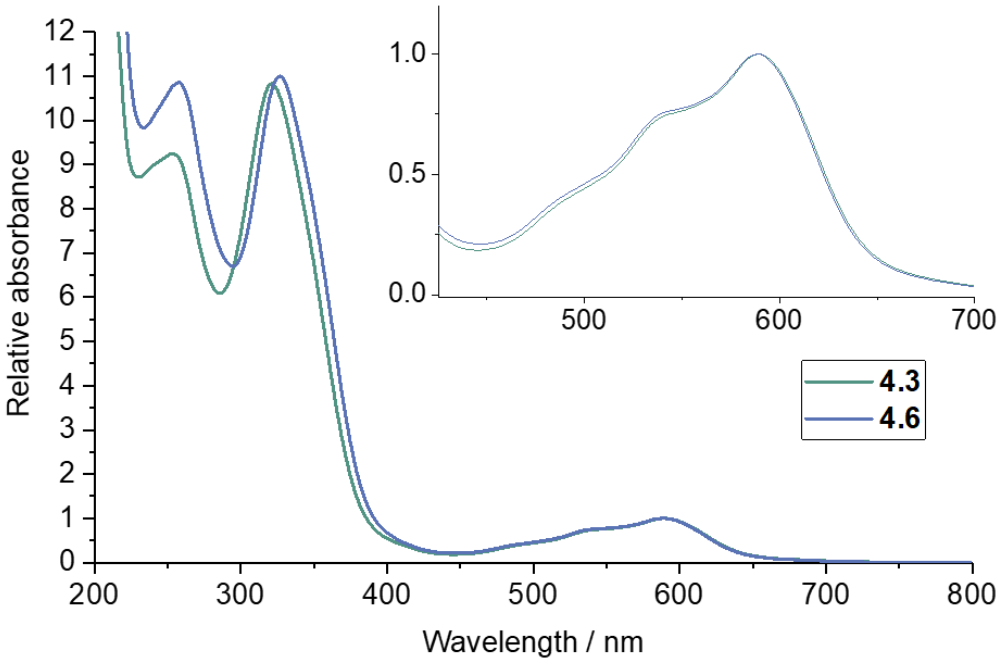


Figure 4.7 | a) UV-vis spectrum of **4.3** (green) and **4.6** (blue) showing the similarity between the two cages. The insert shows a zoom on the MLCT band of both cages.

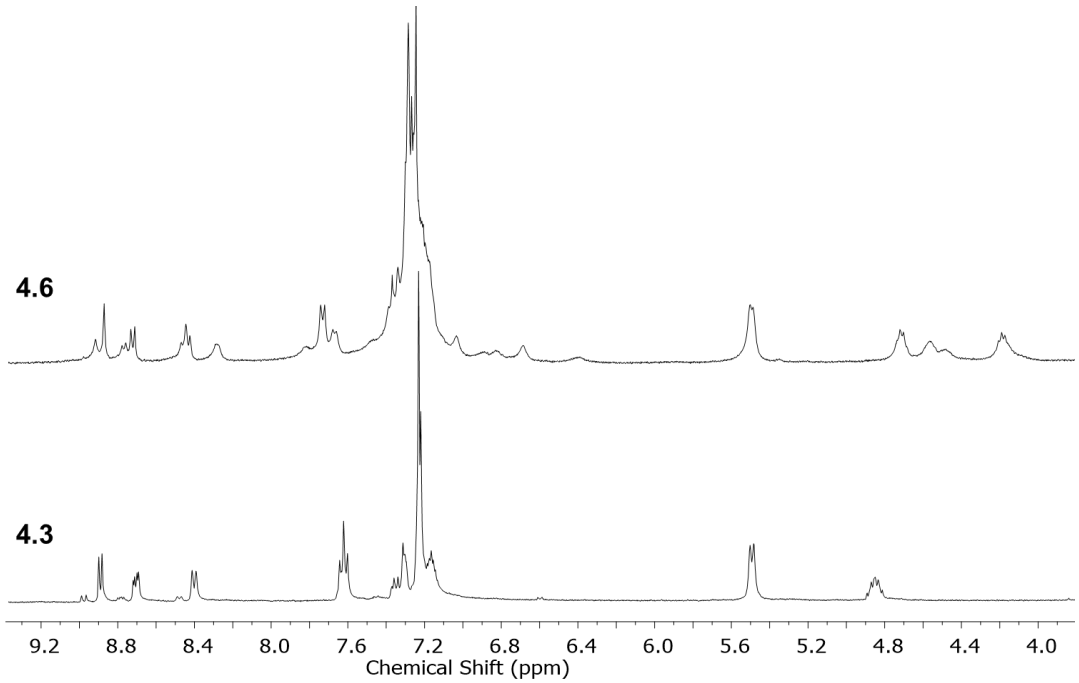


Figure 4.8 | Aromatic region of the ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of **4.6** (top) and $TFA^{-}\subset$**4.3** (bottom) showing the similarity between the two cages.

The PAG was more soluble in CH_3CN/H_2O mixtures than in pure CH_3CN. However, cage **4.3** was found to be unstable in the presence of as little as 5% H_2O in CH_3CN. Decomposition of the cage was evidenced by the release of aniline **4.C** in solution (Figure 4.9). Furthermore the formation of the gel from **4.6** was unsuccessful in the presence of water (1% or more), even after increasing the concentration of cage to 30 mM. As a consequence, the system was limited to pure CH_3CN as a solvent.

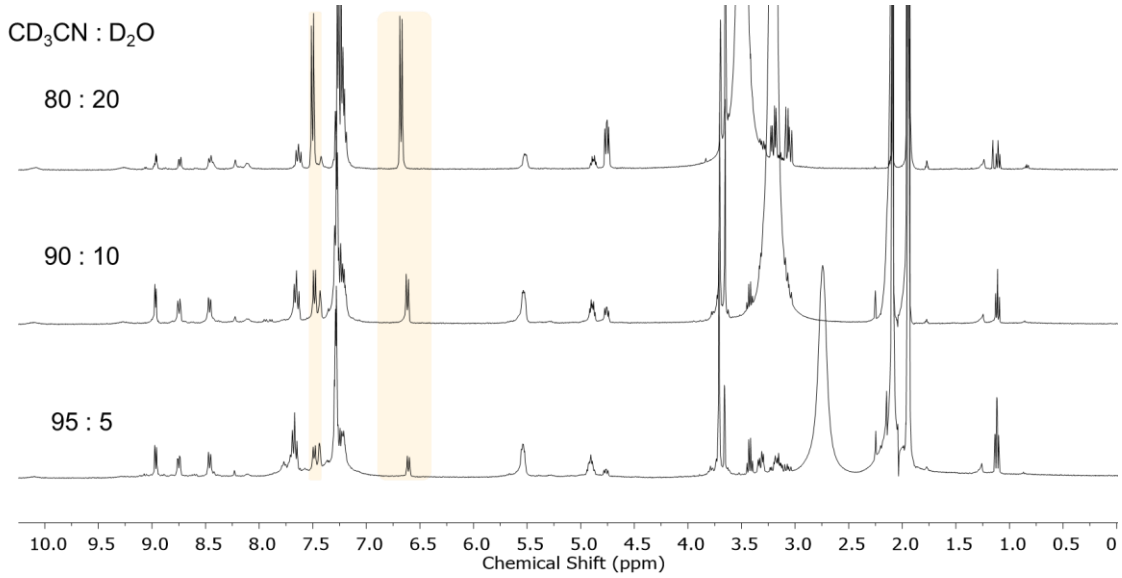


Figure 4.9 | a) ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN:D_2O) of **4.3** in mixtures of CD_3CN and D_2O at different ratios (labelled on the left hand-side). The beige boxes highlight the peaks matching the free aniline **4.C**.

In most reported cases, the PAG was studied in H_2O,[11-13] or mixtures of solvents (including H_2O, $CHCl_3$, CH_3CN, CH_3OH and HFIP).[9, 14-15] Due to the lack of information on the PAG in CH_3CN only, studies were carried out to establish its behaviour in this solvent. As previously reported, the solubility of this compound in CH_3CN was very poor.[16] The maximum solubility was determined by UV-vis spectroscopy. Solutions of **MEH** in CH_3CN (13, 26, 33 and 39 µM) were prepared. The maximum absorbance at 425 nm were plotted against the concentrations, showing a linear relation.

Using Beer-Lambert's law, the molar attenuation coefficient for **MEH** was calculated:

$$\varepsilon = A / \ell \mathcal{C} = 9802\ M^{-1}.cm^{-1}$$

A saturated solution of **MEH** in CH_3CN was prepared by sonicating 5 mg of material in 2 mL of CH_3CN for 1h. The sample was centrifuged and the supernatant was analysed by UV-vis spectroscopy, giving the absorbance at 425 nm A = 0.547. The maximum concentration of **MEH** in pure CH_3CN was then given by:

$$C_{max} = A / (\varepsilon \ell) = 55.8\ \mu M$$

Due to the poor solubility in CH_3CN, ^1H NMR spectra of the PAG were acquired in DMSO-d_6 (Figure 4.10 a). The switching between **MEH** and **SP** happened within a few minutes of exposure to visible light and the reverse reaction happened within an hour in the dark. This observation is however only valid in DMSO as the rate of switching of **MEH** might differ between this solvent and CH_3CN.

Similarly to what was observed in water and solvent mixtures, switching from **MEH** to **SP** was observed in CH_3CN upon irradiation with visible light, as evidenced using UV-visible spectroscopy (Figure 4.10 c). When dissolving **MEH** in CH_3CN in the dark, a yellow solution was obtained. The UV-vis spectrum of **MEH** was dominated by a band at 425 nm. Upon irradiation for 5 min (blue LED, ν = 470 nm approx.), the PAG switched to the **SP** form, leading to the loss of colour of the solution. The band at 425 nm disappeared and was replaced by a band at 301 nm. No bands above 500 nm were observed, indicating the absence of the **ME** transition state and thus the protonation of the PAG (Figure 4.10 c). When isolated from light for 18 h, the metastable **SP** reverted to **MEH** near quantitatively, which could be observed both visually (yellow colour) and by UV-vis spectroscopy.

The change in acidity was investigated in a mixture of CH_3CN and H_2O (95:5) as no values for the pH of the **SP** and **MEH** form can be obtained in pure CH_3CN. As expected, the $pH_{CH3CN/H2O}$ (readout from a pH electrode in the solvent mixture) was dependent on the concentration of

PAG. A drop of $pH_{CH3CN/H2O}$ of 3.2 units caused by the conversion of **MEH** into **SP** was observed upon light irradiation (Figure 4.10 d). Zhang *et al.* reported that in organic solvents, the acidity of the PAG is comparable to that of CH_3SO_3H.[16] Based on the literature and the results observed, we inferred that the change of acidity caused by the conversion of **MEH** to **SP** could be sufficient to impact the gel formation.

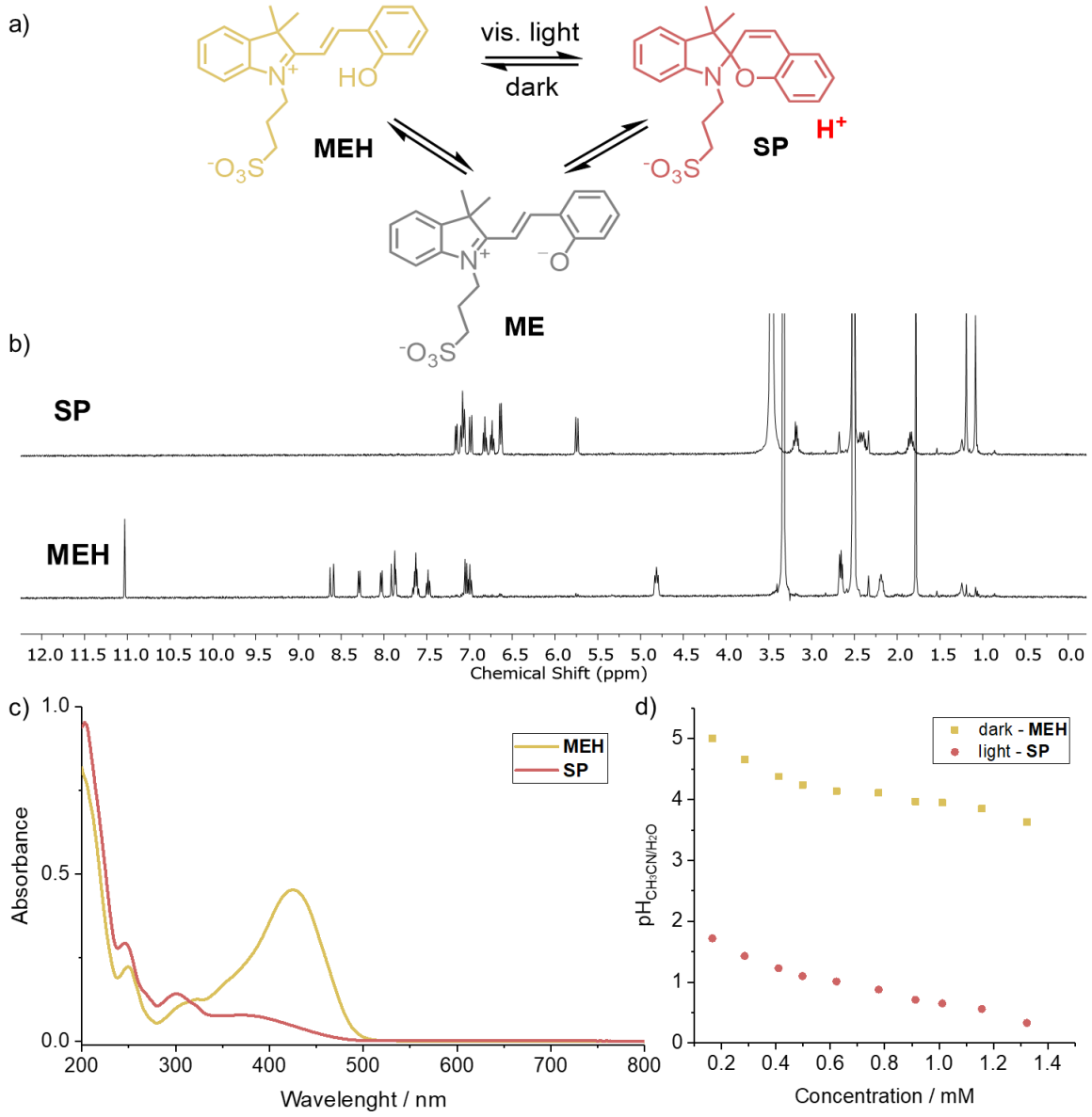


Figure 4.10 | a) Two forms of the PAG, the open ring form in darkness (**MEH**) and the closed ring form once exposed to visible light (**SP**). b) ^{1}H NMR spectrum (400 MHz, 298 K, d_6-DMSO) of **MEH** and **SP**. c) UV-vis spectra of **MEH** (yellow) and **SP** (red) in CH_3CN, showing the typical absorbance bands at 425 nm for **MEH** and 301 nm for **SP**. d) Plot of the pH as a function of the concentration in mM for **MEH** and **SP**.

The transition from **MEH** to **SP** was fast and the irradiation with blue light for 5 min led to the complete conversion of this compound. On the other hand, the kinetic of the reverse process was slow. After irradiation of a 0.05 mM solution of **MEH** in CH_3CN, UV-vis spectra were

acquired every 15 min for 14 h. The gradual increase of the band at 425 nm was indicative of the conversion of **SP** into **MEH** (Figure 4.11 a). Plotting the absorbance at 425 nm allowed us to access information on the kinetic of conversion of **SP** into **MEH** over time. Once fitted to an exponential decay, a tau value of τ = 2.87 ± 0.2 h was obtained (Figure 4.11 b). It is commonly accepted that after five tau, the function reaches a value less than 1% of its starting point and that therefore, the system has reached stability. In this case all of the **SP** was considered to have reverted to **MEH** after 14.5 h.

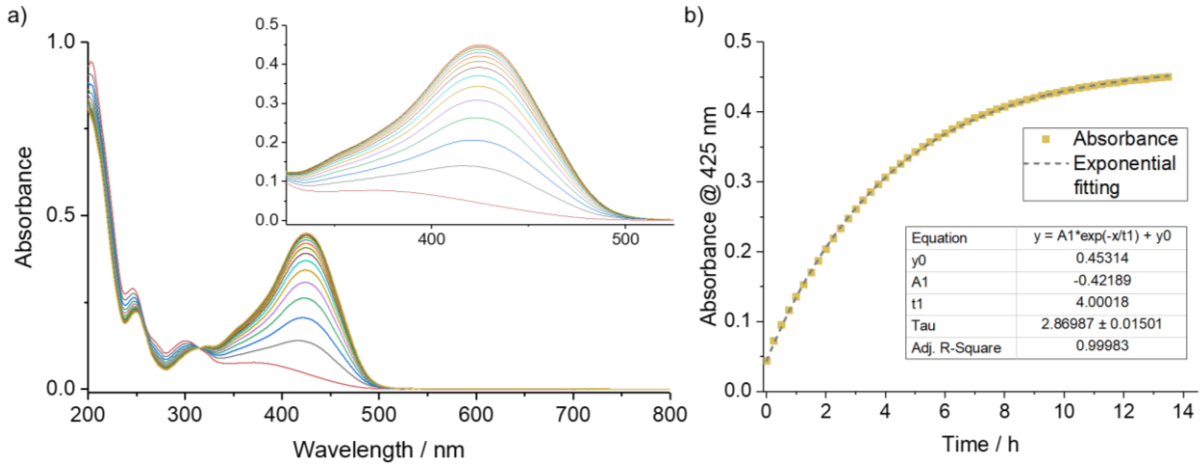


Figure 4.11| a) UV-vis spectra of the PAG in CH_3CN, taken every hour after irradiation of the compound. b) Plot of the absorbance at 425 nm as a function of time for the PAG in CH_3CN along with the exponential fitting.

The compatibility of cage **4.3** and the PAG was probed by NMR and UV-vis measurements. The excitation of the PAG was possible in the presence of the cage, due to the minimal overlap between the absorption bands of the cage (principally the MLCT) and that of the **MEH** (Figure 4.12).

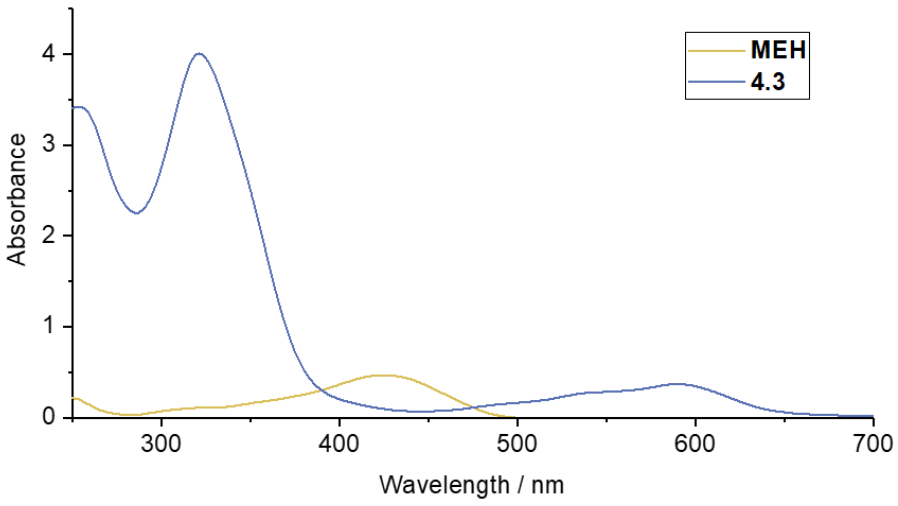


Figure 4.12| UV-vis spectra of **MEH** (0.05 mM) in yellow and **4.3** (0.01 mM) in blue in CH_3CN, showing the typical absorbance bands at 425 nm for **MEH** and 590 nm for **4.3**.

Cage **4.3** (3.7 mg, 0.50 mmol) was dissolved in 0.5 mL of a saturated solution of the PAG in CD_3CN ([**4.3**] = 1 mM and [**PAG**] = 0.05 mM). The sample was kept in the dark for 24 h and the ^1H NMR spectrum of the mixture was recorded. After 5 min of irradiation with blue light, the ^1H NMR spectrum of the mixture was acquired again. This cycle was repeated 4 times, showing no significant decomposition of the cage (Figure 4.13 b). No signal for the PAG was observed as the amount of material present was too small to detect by NMR.

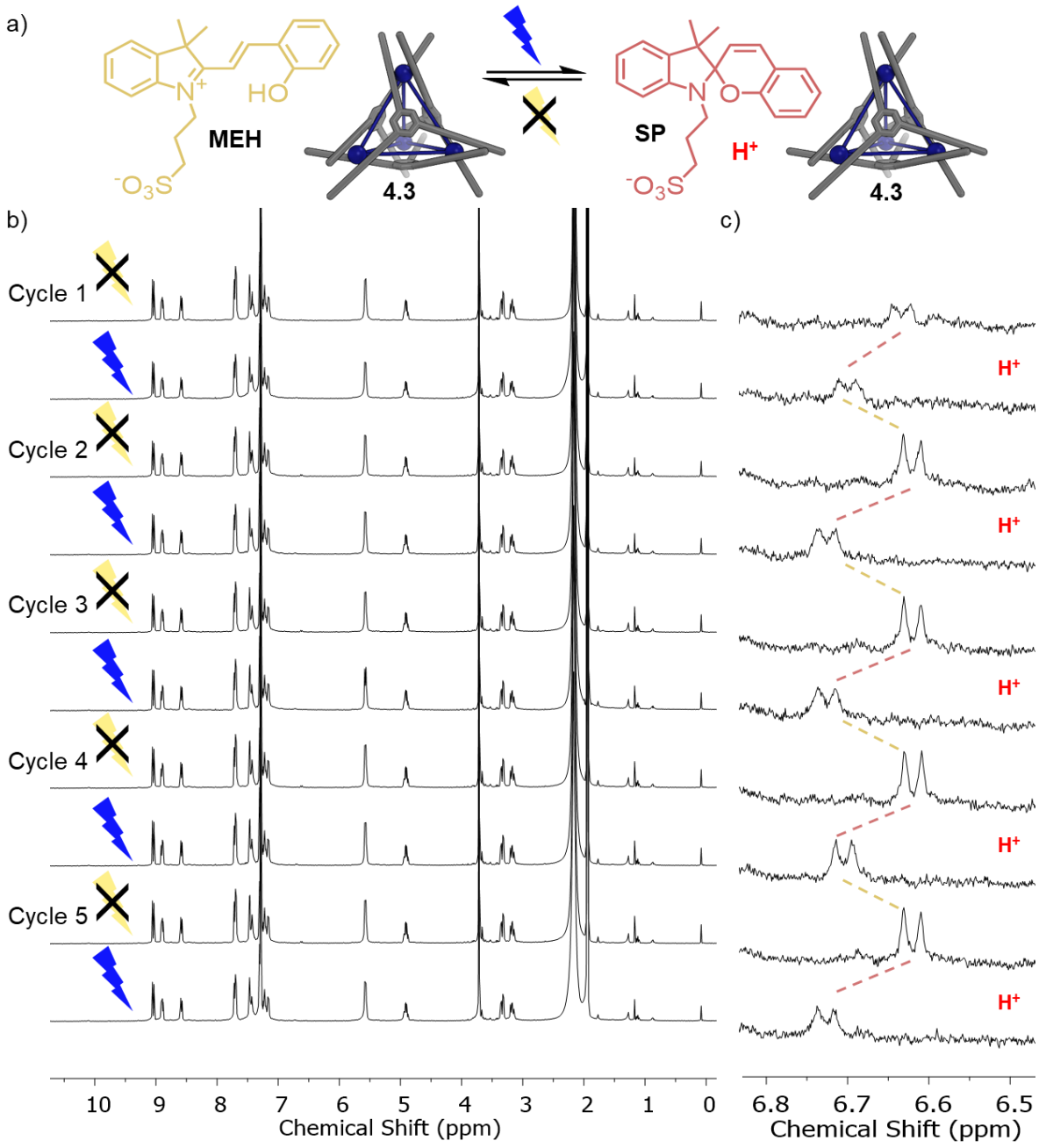


Figure 4.13| a) Composition of the system studied: the PAG (0.05 mM), with its open ring form in the dark (**MEH**) and its closed ring form once exposed to blue light (**SP**) and the cage **4.3** (1 mM). b) ^1H NMR spectra (400 MHz, 298 K, CD_3CN) of **4.3** and the **PAG** over 5 cycles alternating darkness and irradiation showing no degradation of **4.3**. c) Zoom in the 6.8 – 6.5 ppm region of the ^1H NMR spectrum (400 MHz, 298 K, CD_3CN) showing the shift of the free aniline proton peak caused by the change of acidity.

The change of acidity induced by the presence of either **MEH** or **SP** could be observed *via* the shift of the peak at 6.6 ppm corresponding to a small amount of free aniline **4.C** (Figure 4.13 c). This experiment showed that the cage framework was not affected by switching of the PAG and that the PAG could convert between the two forms in the presence of the cage (20 eq.).

A solution of PAG (0.05 mM) and **4.3** (0.01 mM) was used for UV-vis studies. It was necessary to limit the concentration of the cage due to its strong absorbance, leading to saturation of the UV-vis spectrometer detector. As in the NMR studies, the sample was kept in the dark for 24 h and the UV-vis spectrum of the mixture was recorded. After 5 min of irradiation with blue light, the UV-vis spectrum of the mixture was acquired again. The cycle was repeated 4 times. No changes in the intensity of the MLCT band at 590 nm were observed, indicating that the cage was stable in the presence of PAG (5 eq.), while the appearance and disappearance of the band at 420 nm confirmed that the PAG can switch in the presence of the cage (Figure 4.14).

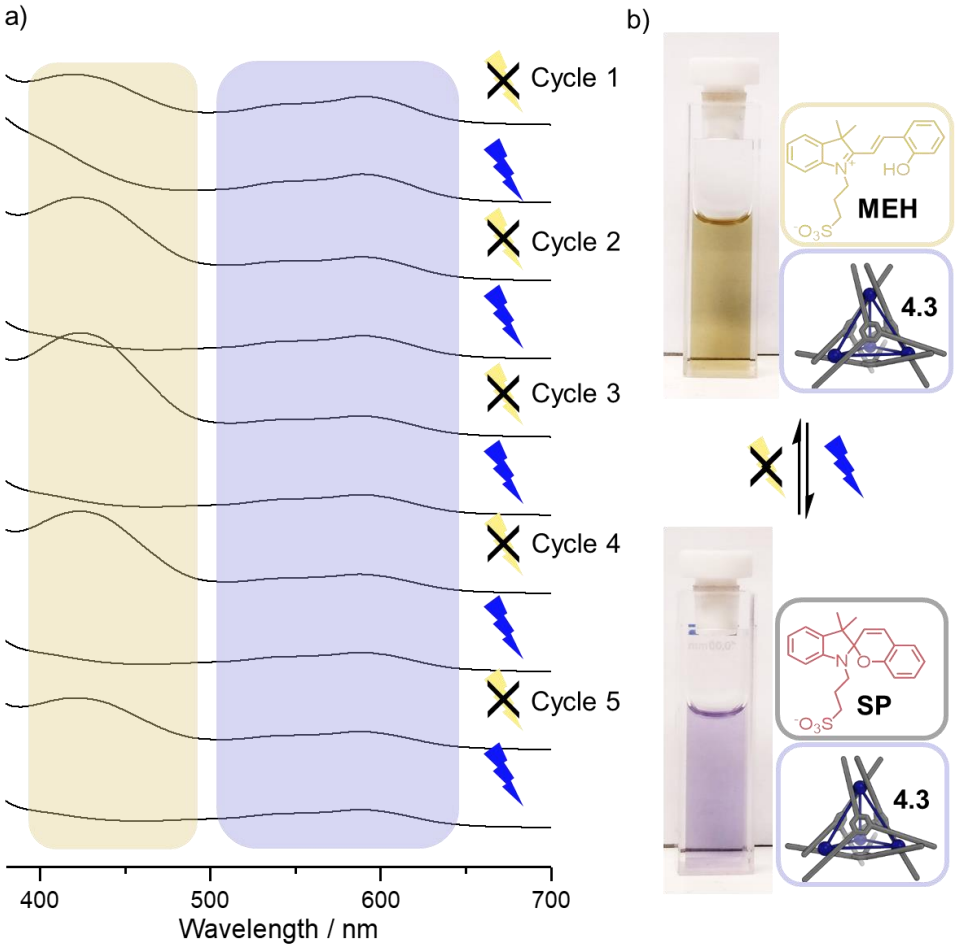


Figure 4.14| a) UV-vis spectra of the PAG (0.05 mM) and **4.3** (0.01 mM) in CH_3CN over 5 cycles alternating darkness (**MEH** form) and irradiation (**SP** form) showing the absence of degradation of **4.3** (blue region) and the appearance and disappearance of the band for **MEH** (yellow region), evidencing the switching between forms. b) Photos of the solutions containing **MEH** and **4.3** in the dark (top) or **SP** and **4.3** after irradiation (bottom).

4.5 Light triggered sol-gel transition of 4.6

Based on the results obtained in Section 4.4, it was inferred that the photoswitching of the PAG in the presence **4.6** should not be impaired and that **4.6** should be stable in the presence of the PAG.

Cage **4.6** (9.2 mg, 0.93 mmol) was dissolved in a saturated solution of PAG (0.056 mM, 100 µl) in the dark. The sample was heated at 50 °C to promote the dissolution of the cage and sonicated for 1 h after which the mixture became viscous. The sample was left in the dark for 18 h after which the gel was formed. The formation of the gel was identified by an inversion test (no movement of the material when the tube was inverted 10 times). Upon exposure to blue light, the sample reverted to solution state within 2 min. After being kept in the dark for 5 h, the gel was formed again. This cycle was performed four times. Significant aggregates were observed in the solution phase, but these did not seem to impair the gel-sol-gel transition. The aggregates were inferred to result from undissolved cage at such high concentrations (10 mM).

When a lower concentration of **4.6** was used (5 mM) with the same concentration of PAG (0.056 mM), the gel initially formed in the dark within 10 h. The gel-sol transition was achieved following 3 min of irradiation and the sol-gel transition happened after 1 h of darkness. The cycle was repeated three times, showing no significant aggregates (Figure 4.15).

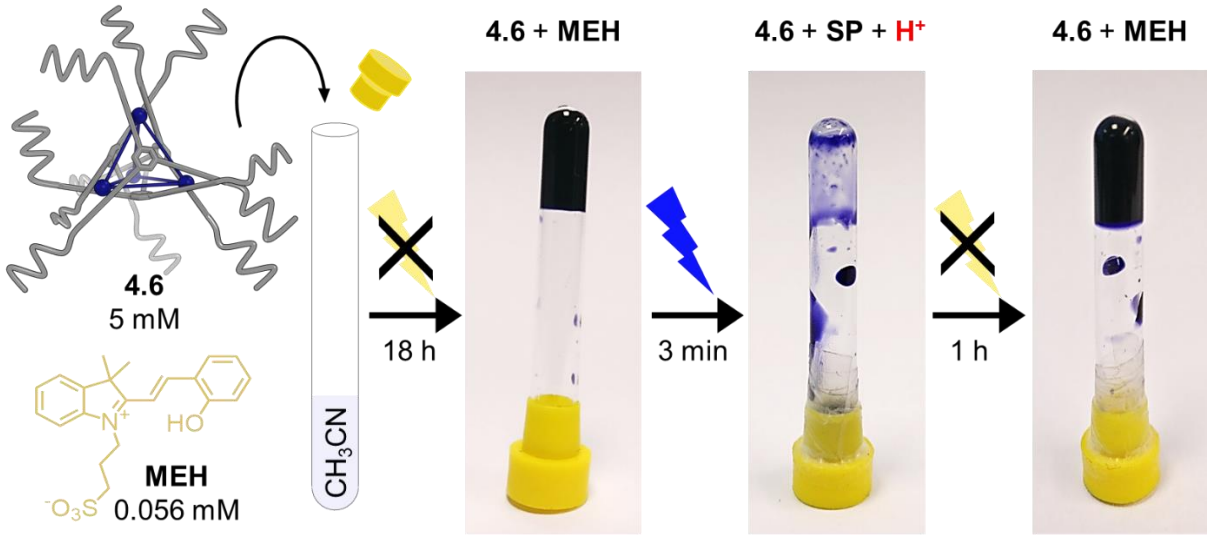


Figure 4.15| Gel-sol-gel transition of a solution of **4.6** (5 mM) and the PAG (0.056 mM) in CH_3CN in the absence of light , after irradiation with blue light for 3 min and after being placed in the dark again for 1 h.

Two solutions of **4.6** (5 mM) were prepared in CH_3CN with (0.056 mM) or without PAG, sonicated for 1 h to promote formation of the gels and kept in the dark for 18 h. Both samples

were simultaneously irradiated with blue light for 10 min. No gel-sol transition was observed in the absence of the PAG, whereas gel-sol transition was observed within 3 min in the presence of the PAG (Figure 4.16). After being placed in the dark for 1 h, the sample underwent sol-gel transition. The gel was allowed to set for 18 h before starting the experiment again. Three more cycles were performed in the same conditions, giving the same outcome each time, ie. gel-sol transition of the sample containing PAG, and absence of changes in the samples without PAG. A video taken during the second cycle of gel-sol transitions can be found in the University repository.

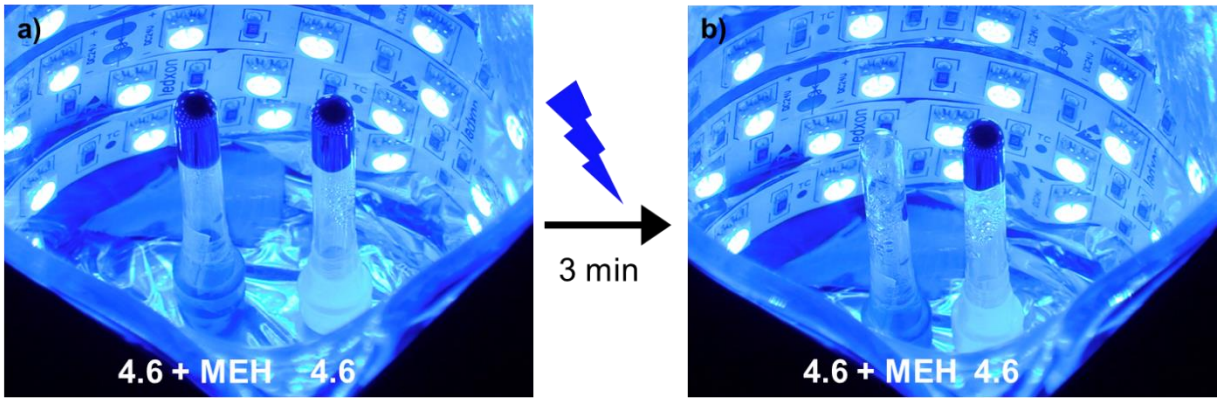


Figure 4.16| Solutions of **4.6** (5 mM) and **MEH** (0.056 mM) in CH_3CN (left) and **4.6** (5 mM) in CH_3CN (right) a) at the beginning of the irradiation with blue light and b) after 3 min of irradiation.

4.6 Conclusions and future work

In this chapter, we have demonstrated the bioconjugation of amino acids and peptides onto metal-organic supramolecular capsules and the changes of properties these can yield on the resulting compounds. High levels of steric hindrance around the vertices interfered with the coordination between the metal centres and the organic ligands formed by condensation between picolinaldehyde and the N-terminus of amino acids, thus preventing the direct assembly of cages from basic amino acids. Addition of a *p*-aminobenzoyl moiety coupled to the N-terminus of amino acids or peptide *via* an amide bond afforded the desired functionalised capsules.

The properties of the biotag (amino acid or peptide) altered the properties of the cage obtained. The chirality of the tag influenced the ratio of the stereoisomers the cage formed, with the degree of induction varying depending on the tag used. When a short peptide gelator was used

to functionalise the cage, the properties of the latter were transferred onto the cage. At high concentration (5 mM and above), the bioconjugated cage formed a gel in acetonitrile.

A spiropyran-based PAG was synthesised and its behaviour in acetonitrile was investigated. The presence of the cage did not impair the switching of the PAG. Furthermore, the cage was stable to the change in acidity triggered by light irradiation, as evidenced by ^{1}H NMR and UV-vis experiments over 5 cycles of switching. When the PAG was added to a solution of the peptide-functionalised cage in the dark, a gel was formed. Upon irradiation with light, a gel-sol transition was achieved. The process was reversible, and multiple cycles of gelation and liquefaction were performed.

Future work will focus on the use of this system for guest encapsulation and release, playing on the difference in kinetics of guest encapsulation in the gel and the solution state. Other potential work could involve the creation of a hydrogel based on a similar cage system in aqueous media by changing the peptide sequence and potentially the face-capping units. The light triggered reversible sol-gel transition of a cage based hydrogel could have potential applications in biomedical fields, for example in drug delivery.

4.7 Experimental section

4.7.1 Synthesis of anilines 4.A – 4.E

General procedure for the synthesis of the methylated amino acid functionalised aniline 4.A – 4.E

The amino acid methyl ester hydrochloride (2.00 mmol, 1.2 eq.), 4-nitrobenzoic acid (279 mg, 1.67 mmol, 1.0 eq.), EDC.HCl (0.627 mg, 3.27 mmol, 2 eq.), DIPEA (0.886 mL, 5.10 mmol, 3 eq.) and HOAt (23 mg, 0.17 mmol, 0.1 eq) were combined in THF (50 mL). The mixture were stirred at room temperature for 18 h. The solvents were removed under reduced pressure and the resulting materials were dissolved in DCM (50 mL). The solutions were washed with water (2 × 50 mL), saturated aqueous NH_4Cl (2 × 50 mL) and saturated aqueous $NaHCO_3$ (2 × 50 mL). The organic phase was dried over $MgSO_4$ and the solvents removed under reduced pressure. The resulting solids were purified by silica gel chromatography (DCM) giving the desired nitrophenyl compound as a yellow solid. The nitrophenyl compound obtained (1.30 mmol) were dissolved in CH_3OH and Pd/C (50 mg) was added. The mixture was stirred

under an atmosphere of H_2 for 3 h. The Pd/C was removed by filtration on celite and the solvents removed under reduced pressure to give the desired anilines as solids.

Methyl (4-aminobenzoyl)glycinate 4.A

Yield over two steps: 77%

^{1}H NMR (400 MHz, 298K, CD_3OD) δ 7.62 (d, *J* = 8.6 Hz, 2H), 6.67 (d, *J* = 8.6 Hz, 2H), 4.07 (s, 2H), 3.74 (s, 3H).

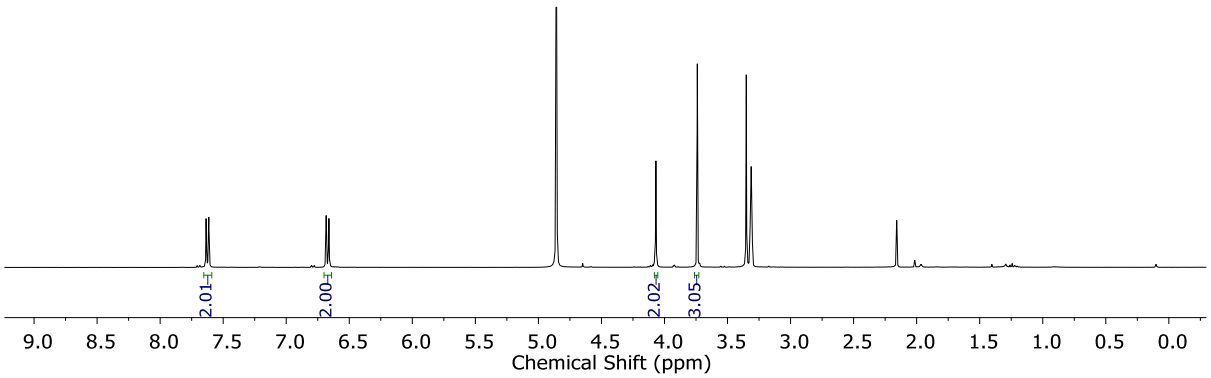


Figure 4.17| ^{1}H NMR (400 MHz, 298K, CD_3OD) of **4.A**.

This data was consistent with previously reported data.[17]

Methyl (4-aminobenzoyl)leucinate 4.B

Yield over two steps: 84%

^{1}H NMR (400 MHz, 298K, $CDCl_3$) δ 7.64 (d, *J* = 8.5 Hz, 2H), 6.70 (d, *J* = 8.5 Hz, 2H), 6.41 (bd, *J* = 8.3 Hz, 1H), 4.84 (td, *J* = 8.4, 4.8 Hz, 1H), 3.76 (s, 3H), 1.84 – 1.55 (m, 4H), 0.98 (t, *J* = 6.0 Hz, 6H).

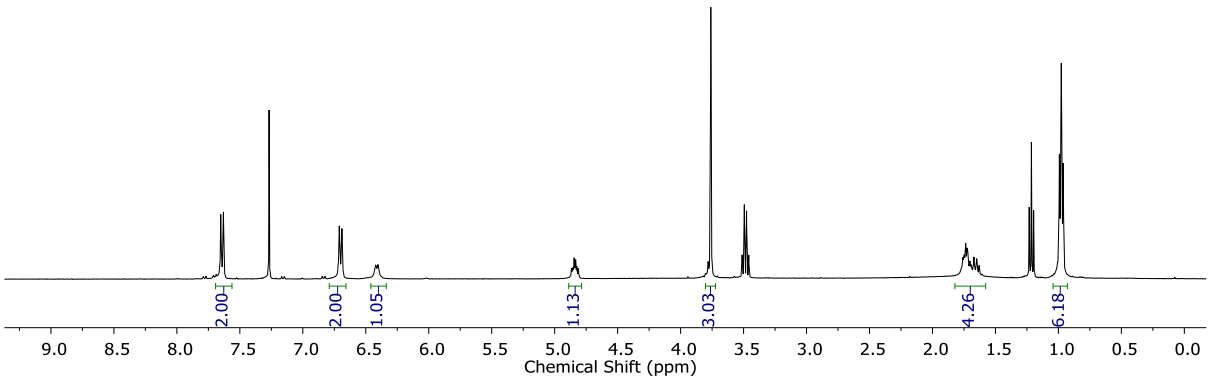


Figure 4.18| ^{1}H NMR (400 MHz, 298K, $CDCl_3$) of **4.B**.

This data was consistent with previously reported data.[18]

Methyl (4-aminobenzoyl)phenylalaninate 4.C

Yield over two steps: 89%

^{1}H NMR (400 MHz, 298K, $CDCl_3$) δ 7.56 (d, *J* = 8.6 Hz, 2H), 7.32 – 7.20 (m, 3H), 7.17 – 7.04 (m, 2H), 6.64 (d, *J* = 8.6 Hz, 2H), 6.40 (bd, *J* = 7.7 Hz, 1H), 5.14 – 4.93 (m, 1H), 3.95 (bs, 1H), 3.74 (s, 3H), 3.36 – 3.13 (m, 2H).

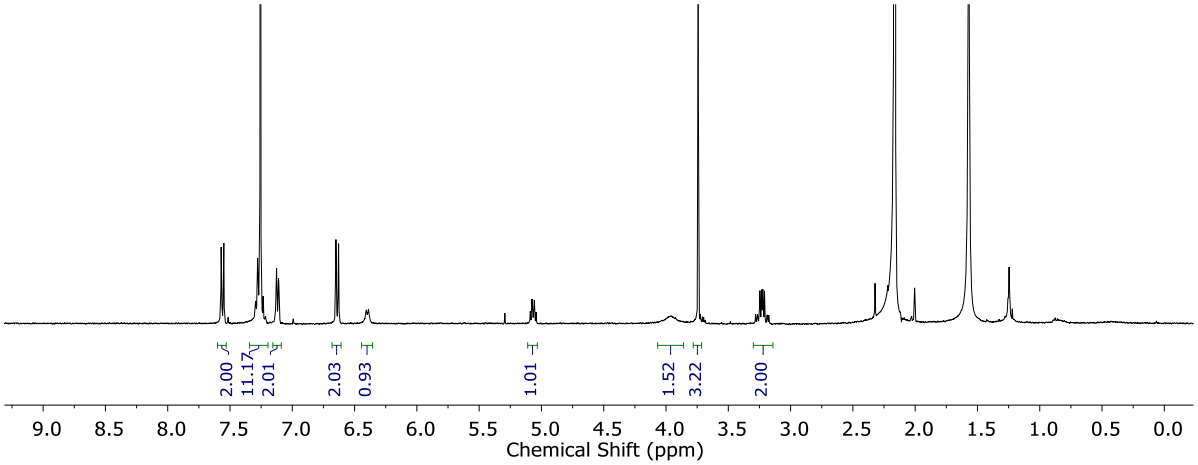
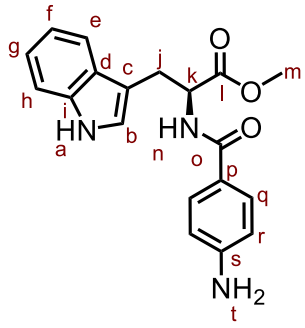


Figure 4.19 | ^{1}H NMR (400 MHz, 298K, $CDCl_3$) of **4.C**.

This data was consistent with previously reported data.[19]

Methyl (4-aminobenzoyl)tryptophanate 4.D and 4.E



4.D and **4.E** are enantiomers and therefore have identical spectra. Only **4.D** has been fully characterised.

Yield over two steps: 82% (**4.D**) and 76% (**4.E**)

^{1}H NMR (400 MHz, 298K, $CDCl_3$) δ 8.21 (s, 1H, H_a), 7.55 (d, *J* = 8.4 Hz, 1H, H_e), 7.51 (d, *J* = 8.6 Hz, 1H, H_q), 7.34 (d, *J* = 8.1 Hz, 1H, H_h), 7.17 (t, *J* = 7.5 Hz, 1H, H_f), 7.07 (t, *J* = 7.5 Hz, 1H, H_g), 6.99 (d, *J* = 2.4 Hz, 1H, H_b), 6.58 (d, *J* = 8.6 Hz, 2H, H_r), 6.52 (d, *J* = 7.7 Hz, 1H, H_n), 5.17 – 5.08 (m, 1H, H_k), 3.95 (bs, 2H, H_t), 3.70 (s, 3H, H_m), 3.42 (d, *J* = 5.2 Hz, 2H, H_j). **^{13}C NMR** (101 MHz, 298K, $CDCl_3$) δ 172.3, 166.3, 149.4, 135.8, 128.6, 127.3, 123.0, 122.5, 121.9, 119.3, 118.4, 113.7, 110.9, 109.9, 52.9, 52.0, 27.4. **HRMS:** *m/z* calculated for [**4.D** + Na^+] = 360.1324, observed = 360.1292 and for [**4.D**.HCl] = 373.1193, observed = 373.1159

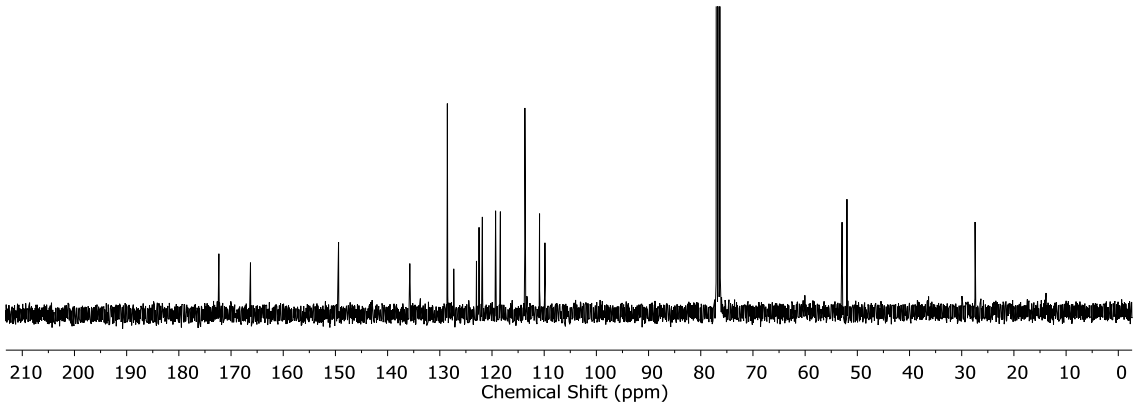


Figure 4.20| ^{13}C NMR (101 MHz, 298K, $CDCl_3$) of **4.D**.

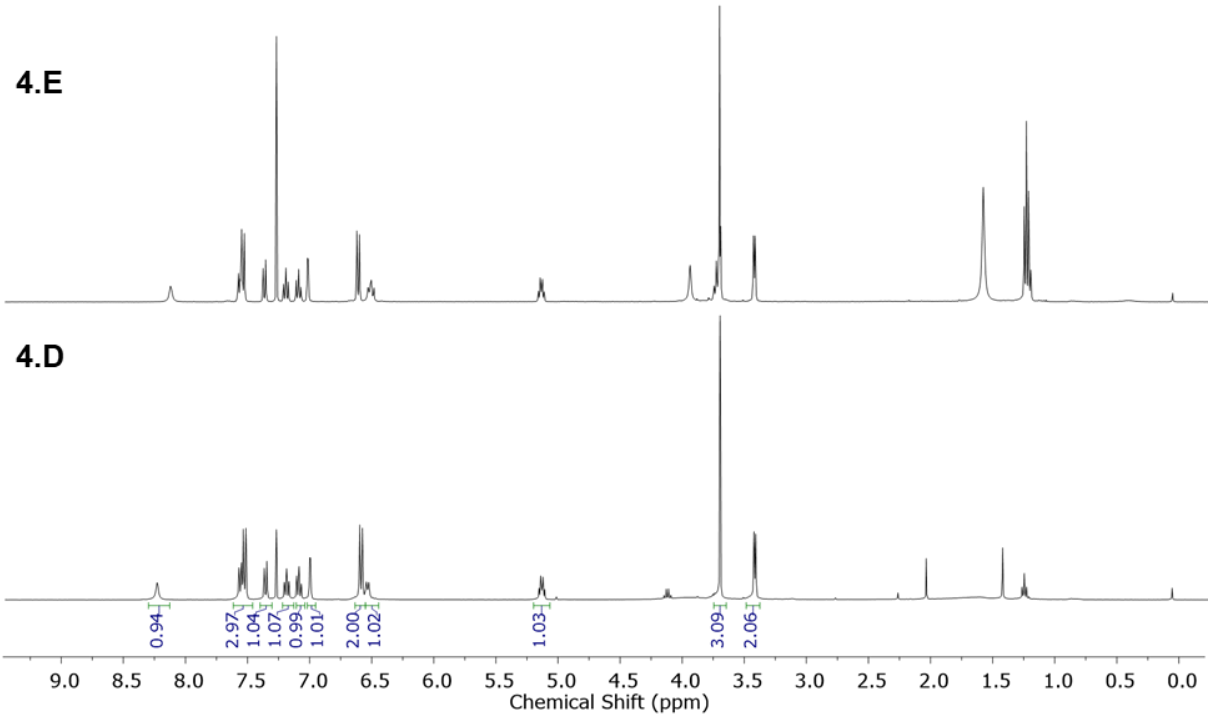


Figure 4.21| ^{1}H NMR (400 MHz, 298K, $CDCl_3$) of **4.E** (top) and **4.D** (bottom).The small difference in chemical shift for the indole and imine protons of the two compounds was attributed to small changes in water content.

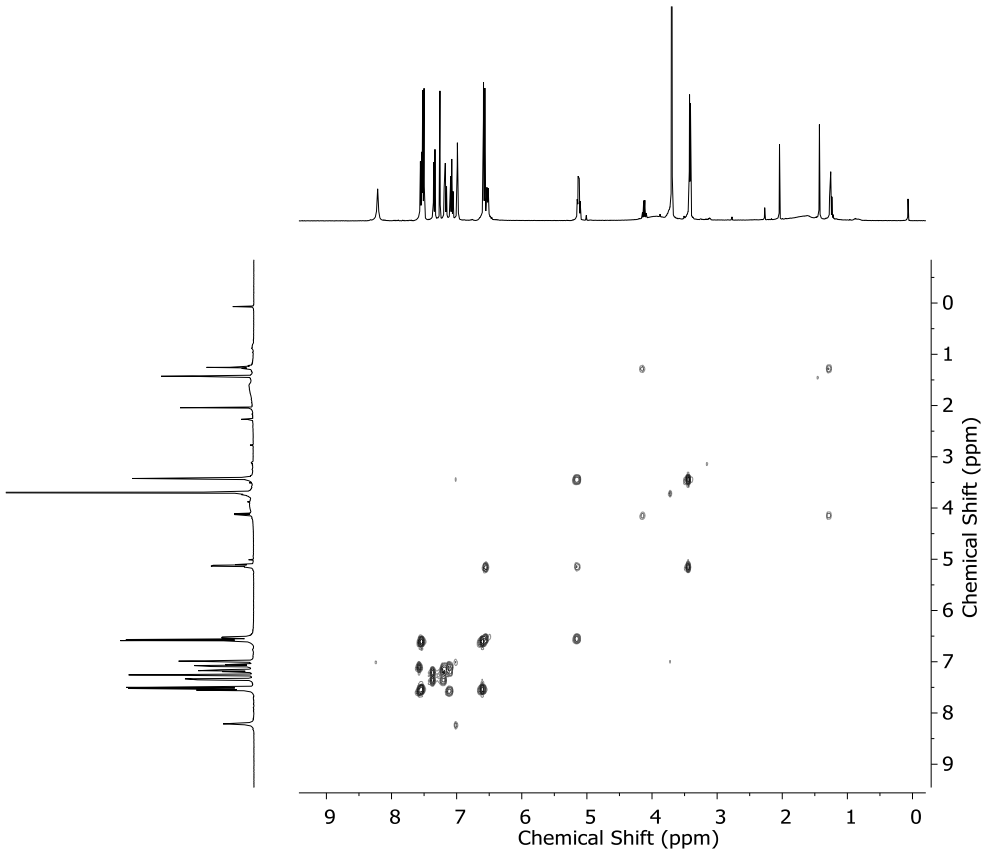


Figure 4.22| ^{1}H-^{1}H COSY spectrum (400 MHz, 298 K, $CDCl_3$) of **4.D.**

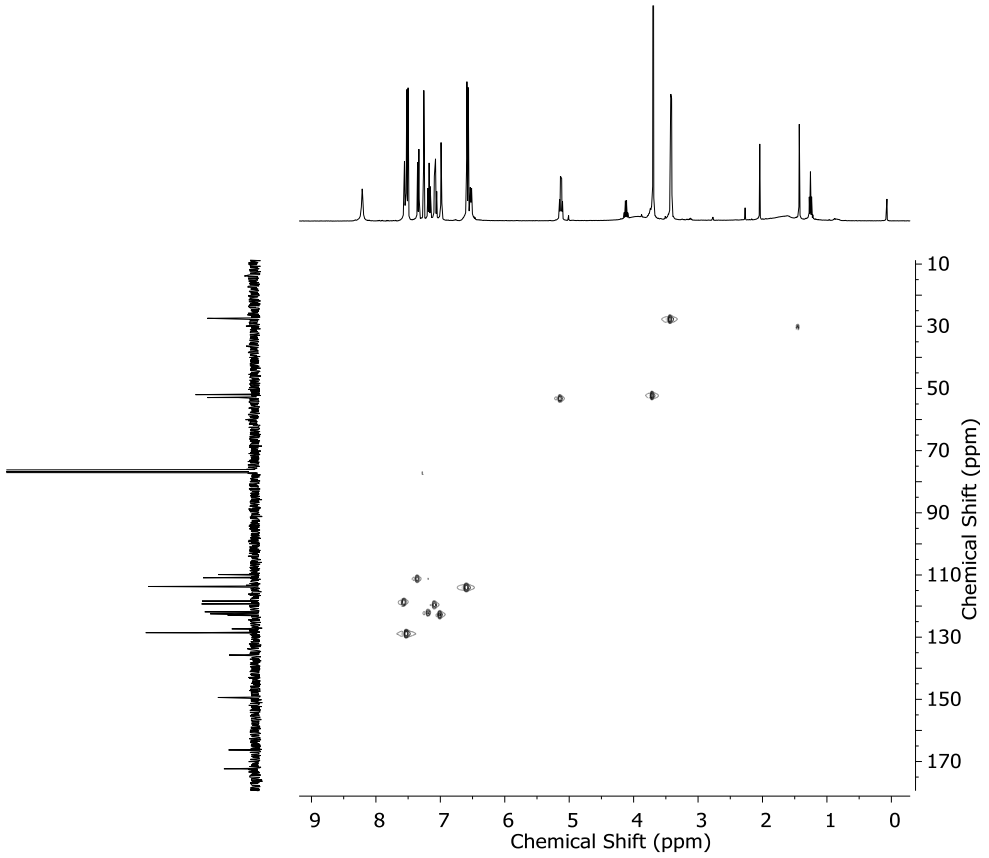
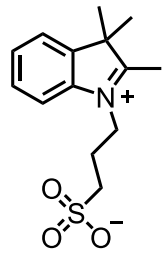


Figure 4.23| ^{1}H-^{13}C HSQC spectrum (400 MHz, 298 K, $CDCl_3$) of **4.D.**

4.7.2 Synthesis of PAG

MEH was synthesized from an adapted literature procedure.[10]

3-(2,3,3-trimethyl-3H-indol-1-ium-1-yl)propane-1-sulfonate 4.F.



1,3-propane sultone (1.26 g, 10.0 mmol, 1 eq.) was combined with 2,3,3-trimethylindolenine (1.65 g, 10.0 mmol, 1 eq.) in a round bottom flask under N_2. The mixture was heated at 90 °C for 4 h. After cooling to room temperature, the solid was isolated by filtration, washed with Et_2O and dried under vacuum to give 3-(2,3,3-trimethyl-3H-indol-1-ium-1-yl)propane-1-sulfonate as a purple glass-like solid (2.52 g, 9.0 mmol, η = 86.5%). The obtained compound was used without further purification.

^{1}H NMR (400 MHz, 298K, D_2O) δ 7.72 (t, *J* = 6.6, 1H), 7.63 (t, *J* = 5.8, 1H), 7.56 – 7.49 (m, 2H), 4.54 (t, *J* = 7.5, 2H), 3.00 (t, *J* = 7.5, 2H), 2.35 – 2.21 (m, 2H), 1.46 (s, 6H).

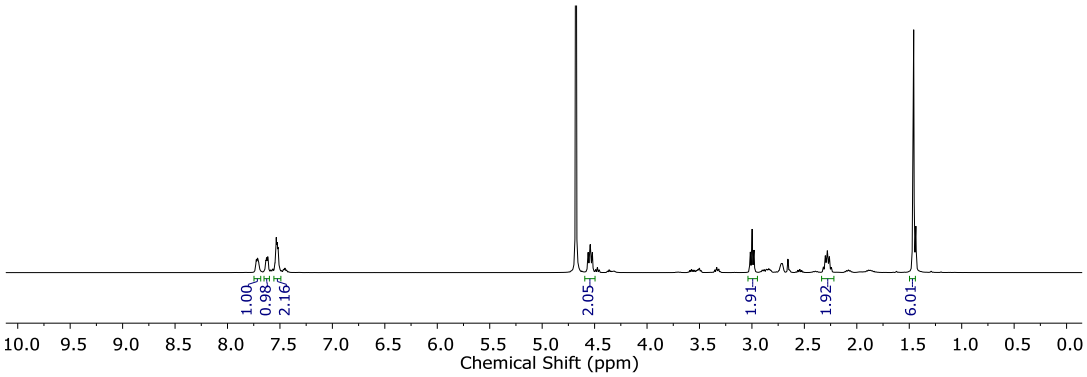
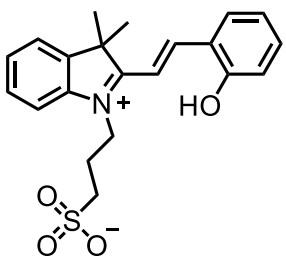


Figure 4.24| ^{1}H NMR (400 MHz, 298K, D_2O) of **4.F**.

(E)-3-(2-(2-hydroxystyryl)-3,3-dimethyl-3H-indol-1-ium-1-yl)propane-1-sulfonate MEH.



3-(2,3,3-trimethyl-3H-indol-1-ium-1-yl)propane-1-sulfonate (504 mg, 1.8 mmol, 1 eq.) was dissolved in 10 mL of EtOH. 2-Hydroxybenzaldehyde (210 µL, 1.95 mmol, 1.1 eq.) was added to the solution and the mixture was refluxed for 18 h under N_2. After cooling to room temperature, the solid was isolated by filtration, washed with Et_2O and dried under vacuum to give **SP** as an orange solid (0.34 g, 0.88 mmol, η = 49 %).

^{1}H NMR (400 MHz, 298K, DMSO-d_6) δ 11.08 (s, 1H), 8.60 (d, *J* = 16.3 Hz, 1H), 8.26 (d, *J* = 7.4 Hz, 1H), 8.01 (d, *J* = 7.4 Hz, 1H), 7.91 – 7.80 (m, 2H), 7.67 – 7.57 (m, 2H), 7.47 (t, *J* = 7.9, 1H), 7.06 – 6.92 (m, 2H), 4.79 (t, *J* = 7.9 Hz, 2H), 2.66 (t, *J* = 7.1, 2H), 2.21 – 2.14 (m, 2H), 1.77 (s, 6H).

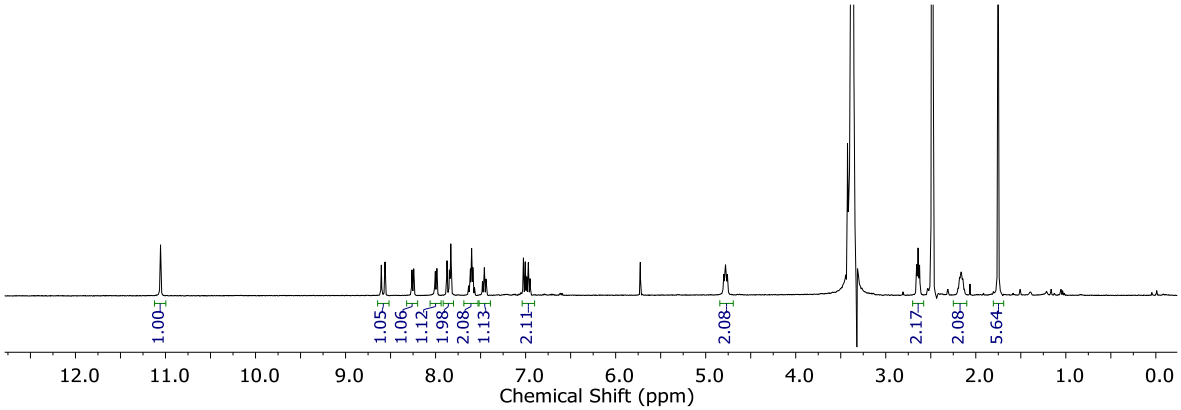


Figure 4.25 | ^{1}H NMR (400 MHz, 298K, DMSO-d_6) of **MEH.**

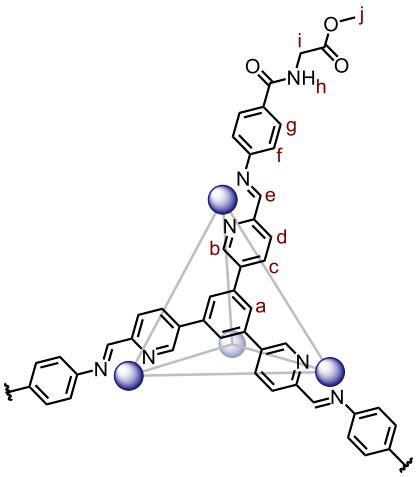
This data was consistent with previously reported data.[10]

4.7.3 Synthesis and characterisation of cages 4.1 – 4.6

General procedure for the synthesis of cages 4.1 – 4.6.

5,5',5"-(benzene-1,3,5-triyl)tripicolinaldehyde **3.A** (3.6 mg, 0.0092 mmol, 1.0 eq.) and aniline **4.A** − **4.E** or **3.C** (0.0275 mmol, 3.0 eq.) were dissolved in 2.0 mL of CH_3CN in a sealed 5 mL schlenk flask. Oxygen was removed by freeze-pumped-thawing with N_2 three times after which $Fe(NTf_2)_2$ (6.3 mg, 0.0023 mmol, 1.0 eq.) was added. The solution was freeze-pumped-thawed an additional two times. The solution was heated at 50 °C for 18 h. The dark blue solution was then cooled and concentrated under a flow of nitrogen. Addition of 30 mL of diethyl ether precipitated the compound into a dark blue solid. The solid was separated by centrifugation and washed with diethyl ether (2 × 30 mL). The solid was then dried under a flow of nitrogen.

Cage 4.1



^{1}H NMR (400 MHz, 298K, CD_3CN) δ 9.11 (s, 12H, H_e), 8.93 (d, J = 8.4 Hz, 12H, H_d), 8.61 (d, J = 8.4 Hz, 12H, H_c), 7.83 (d, J = 8.3 Hz, 24H, H_g), 7.55 (bt, J = 5.7 Hz, 12H, H_h), 7.49 (s, 12H, H_a), 7.21 (s, 12H, H_b), 5.68 (d, J = 8.3 Hz, 24H, H_f), 4.15 (d, J = 6.1 Hz, 24H, H_i), 3.74 (s, 36H, H_j). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 176.8, 171.3, 166.8, 159.5, 153.4, 152.7, 138.9, 137.5, 136.9, 135.4, 133.9, 129.8, 127.0, 122.7, 122.0, 52.8, 42.2. **LR-ESI-MS**: [charge, calculated mass]: *m/z* = 1238.5 [**4.1**$Cl(NTf_2)_3^{4+}$, 1238.5], 934.9

[**4.1**Cl$(NTf_2)_2^{5+}$, 934.8], 732.4 [**4.1**Cl$(NTf_2)^{6+}$, 732.3], 587.7[**4.1**Cl^{7+}, 587.7]. Note: The cage picked up a Cl^- anion, probably due to the mass spectrometer being dirty. **HR-ESI-MS**: *m/z* calculated for **4.1**$(NTf_2)_3^{5+}$ = 938.9697, observed = 938.9693.

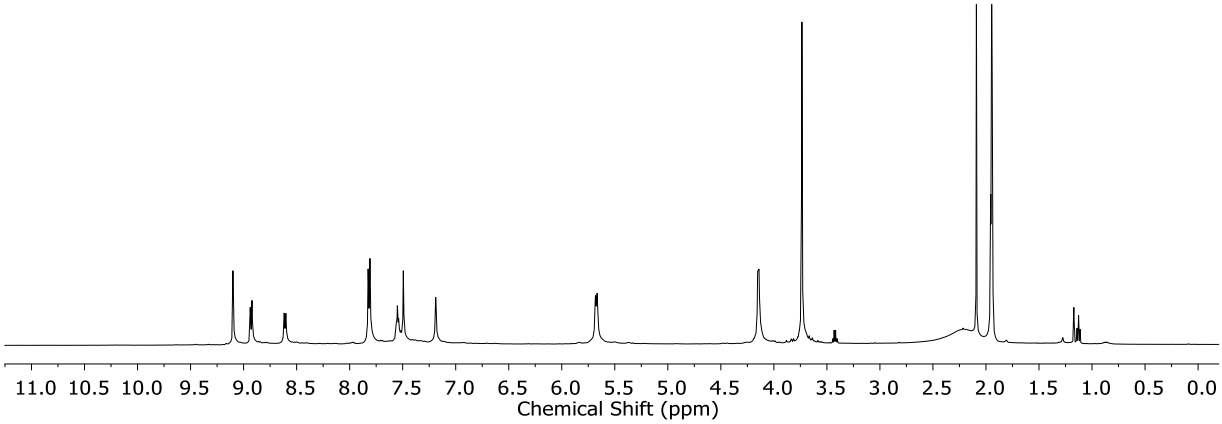


Figure 4.26 | ^{1}H NMR (400 MHz, 298K, CD_3CN) of **4.1.**

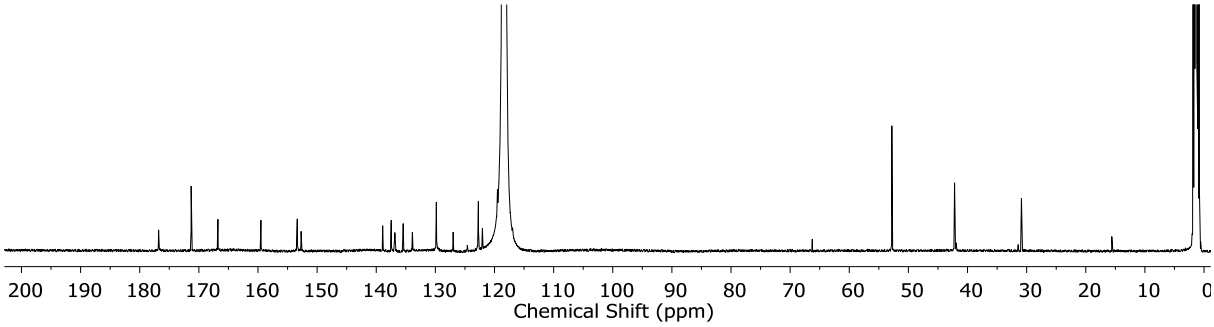


Figure 4.27 | ^{13}C NMR (126 MHz, 298K, CD_3CN) of **4.1.**

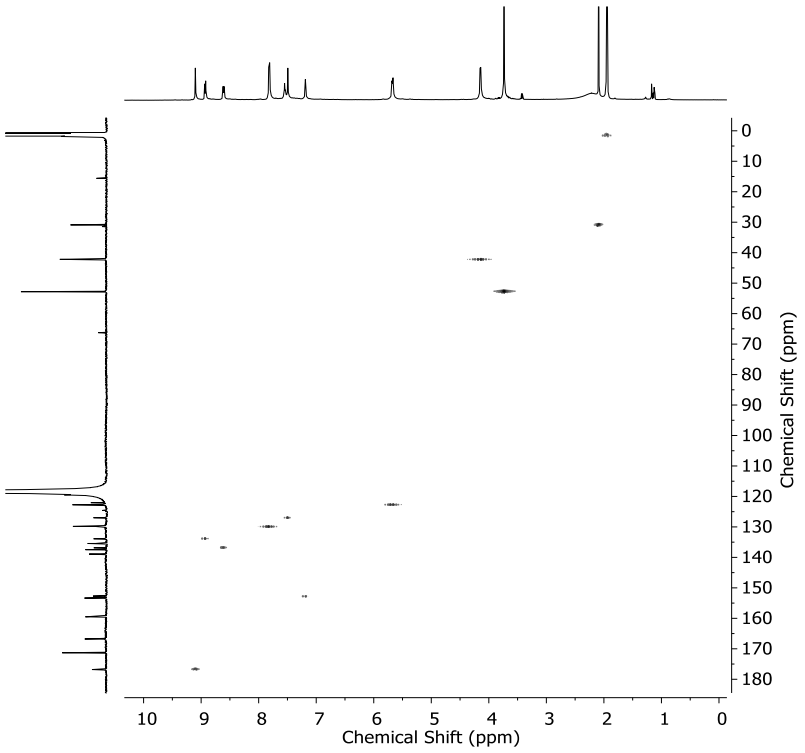
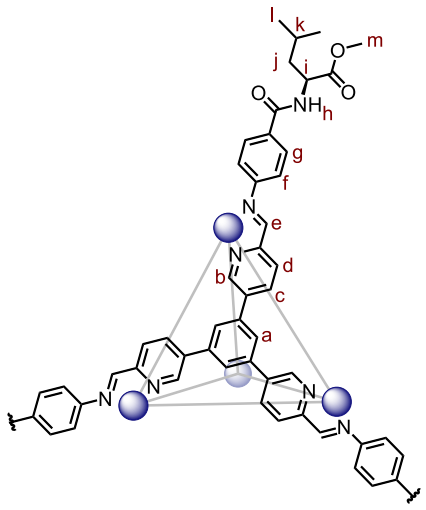


Figure 4.28 | ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, CD_3CN) of **4.1**.

Cage 4.2



^{1}H NMR (400 MHz, 298K, CD_3CN) δ 9.10 (s, 12H, H_e), 9.10 (s, 12H, H_e), 8.94 (d, *J* = 8.4 Hz, 24H, H_d), 8.61 (d, *J* = 8.4 Hz, 24H, H_c), 7.87 – 7.76 (m, 48H, H_g), 7.50 (s, 24H, H_a), 7.41 – 7.34 (m, 24H, H_b), 7.19 (bs, 24H, H_h), 5.71 – 5.59 (m, 48H, H_f), 4.73 – 4.68 (m, 24H, H_i), 3.73 (s, 36H, H_m), 3.72 (s, 36H, H_m), 1.88 – 1.67 (m, 72H, $H_{k,j}$), 1.00 (ddd, *J* = 9.5, 6.1, 3.3 Hz, 144H, H_l). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 176.8, 174.1, 166.7, 159.5, 153.3, 152.8, 138.9, 137.5, 136.9, 135.5, 133.8, 130.0, 127.0, 122.6, 122.1, 53.3 – 51.8 (m), 40.9, 25.8, 23.2, 21.8. **LR-ESI-MS**: [charge, calculated mass]: *m/z* = 1468.2 [**4.2**$(NTf_2)_4^{4+}$, 1467.9], 1118.6 [**4.2**$(NTf_2)_3^{5+}$, 1118.3], 885.4 [**4.2**$(NTf_2)_2^{6+}$, 885.3], 718.8 [**4.2**$(NTf_2)^{7+}$, 718.8]. **HR-ESI-MS**: *m/z* calculated for **4.2**$(NTf_2)_4^{4+}$ = 1467.3746, observed = 1467.3689.

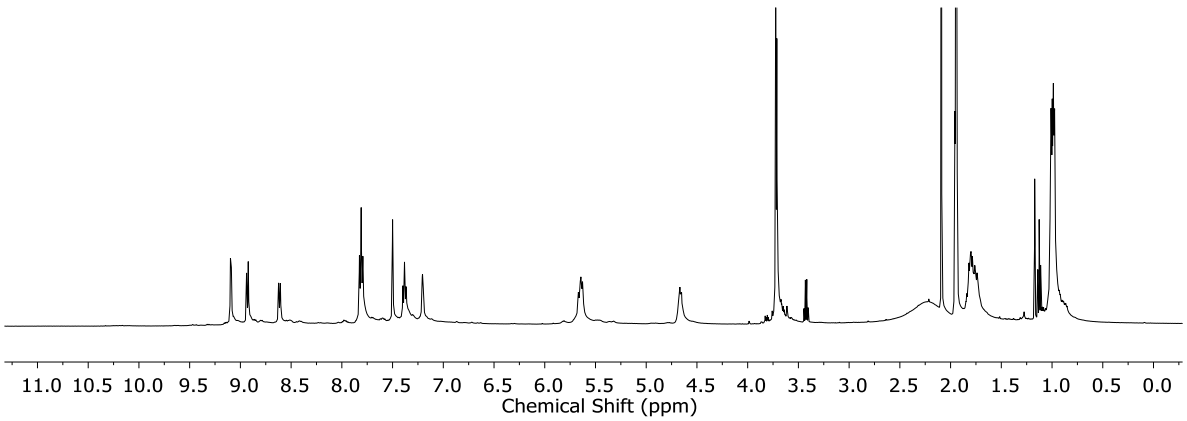


Figure 4.29 | ^{1}H NMR (400 MHz, 298K, CD_3CN) of **4.2.**

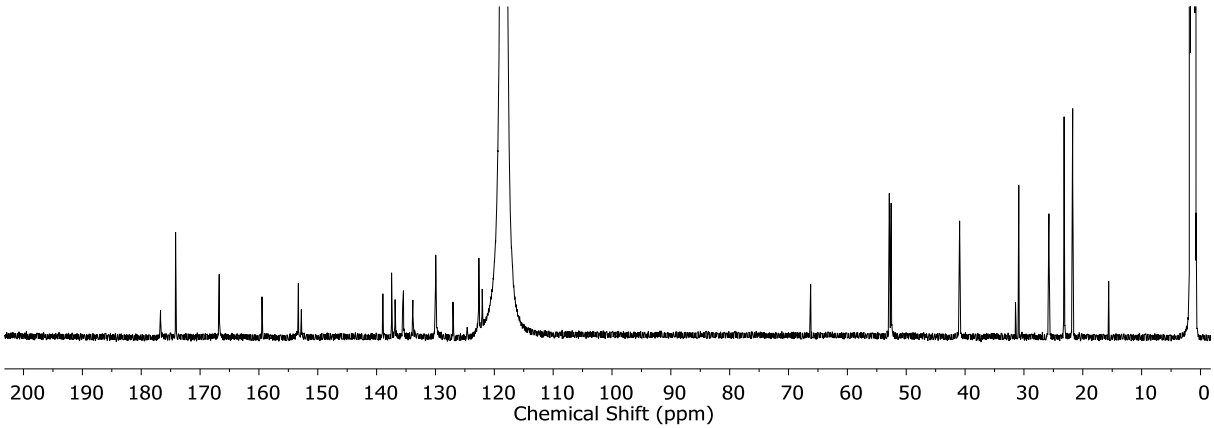


Figure 4.30 | ^{13}C NMR (126 MHz, 298K, CD_3CN) of **4.2.**

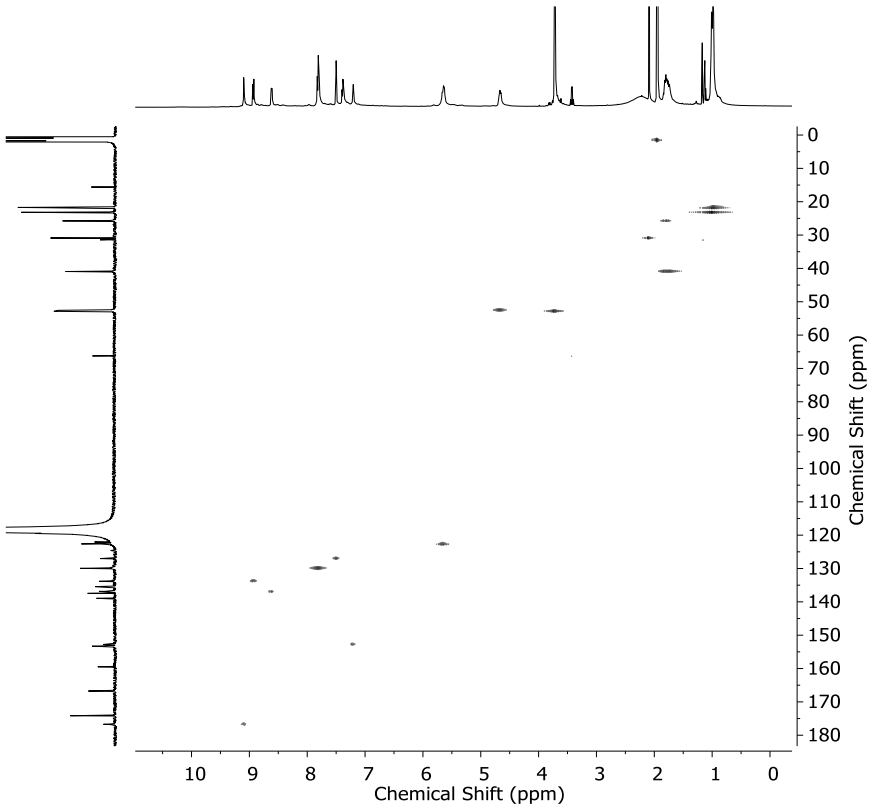
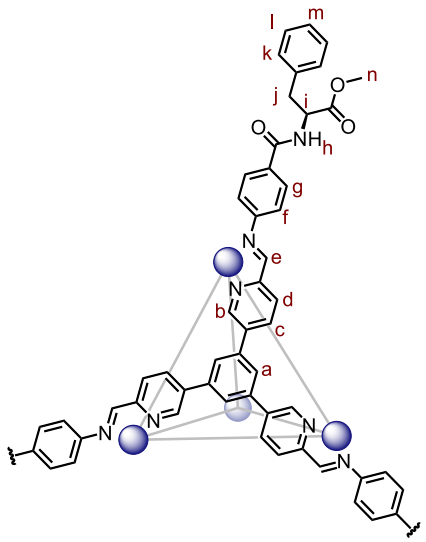


Figure 4.31 | ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, CD_3CN) of **4.2**.

Cage 4.3



^{1}H NMR (400 MHz, 298K, CD_3CN) δ 9.07 (s, 12H, H_e), 9.04 (s, 12H, H_e), 8.91 (d, *J* = 4.2 Hz, 12H, H_d), 8.89 (d, *J* = 4.3 Hz, 12H, H_d), 8.58 (d, *J* = 8.4 Hz, 24H, H_c), 7.75 – 7.67 (m, 48H, H_g), 7.47 (d, *J* = 2.4 Hz, 24H, H_h), 7.44 – 7.14 (m, 180H, $H_{a,b,k,l,m}$), 5.58 (d, *J* = 8.1 Hz, 48H, H_f), 5.01 – 4.85 (m, 24H, H_i), 3.73 (s, 72H, H_n), 3.41 – 3.10 (m, 48H, H_j). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 176.7, 172.9, 166.4, 159.4, 153.3, 152.8, 139.0, 138.1, 137.4, 136.8, 135.4, 133.8, 130.3, 129.9, 129.5, 127.9, 127.0, 122.6, 122.1, 55.4, 53.0, 38.0. **LR-ESI-MS**: [charge, calculated mass]: *m/z* = 1570.3 [**4.3**$(NTf_2)_4^{4+}$, 1569.9], 1200.2 [**4.3**$(NTf_2)_3^{5+}$, 1199.9], 953.5 [**4.3**$(NTf_2)_2^{6+}$, 953.3], 777.2 [**4.3**$(NTf_2)^{7+}$, 777.1]. **HR-ESI-MS**: *m/z* calculated for **4.3**$(NTf_2)_3^{5+}$ = 1200.0794, observed = 1200.0829.

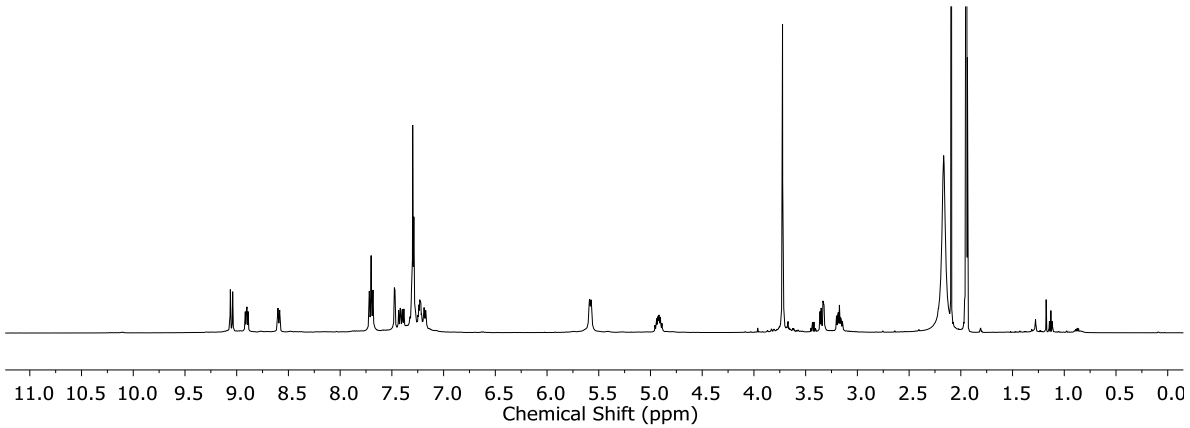


Figure 4.32| ^{1}H NMR (400 MHz, 298K, CD_3CN) of **4.3.**

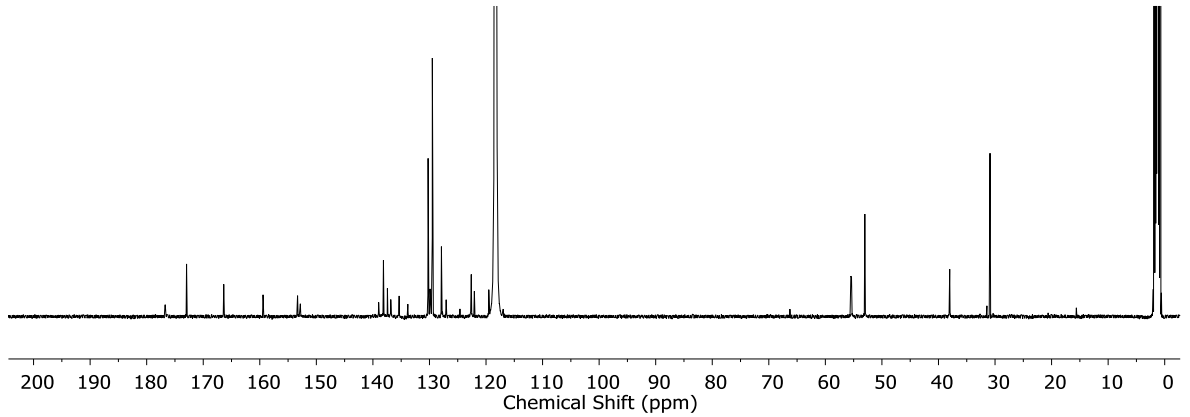


Figure 4.33| ^{13}C NMR (126 MHz, 298K, CD_3CN) of **4.3.**

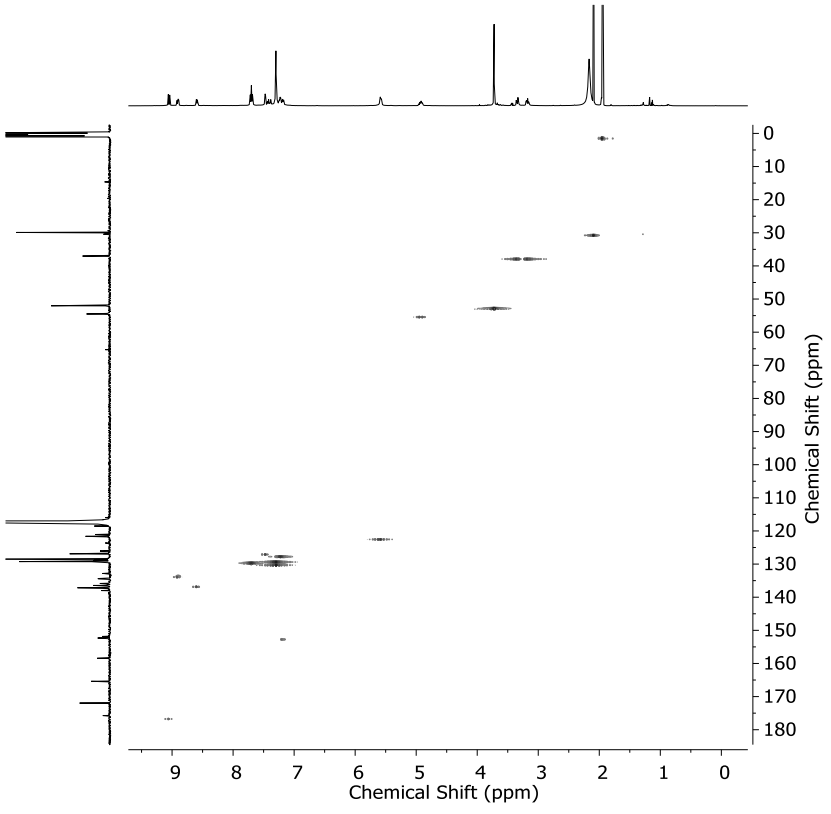
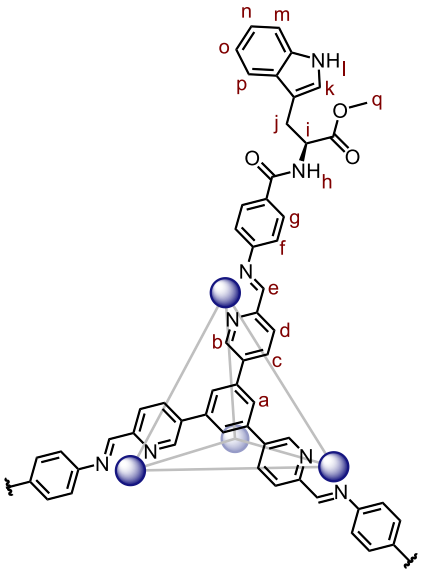


Figure 4.34| ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, CD_3CN) of **4.3**.

Cage 4.4



^{1}H NMR (400 MHz, 298K, CD_3CN) δ 9.48 (s, 12H, H_l), 9.41 (s, 12H, H_l), 9.04 (s, 12H, H_e), 9.01 (s, 12H, H_e), 8.94 – 8.83 (m, 24H, H_d), 8.63 – 8.52 (m, 24H, H_c), 7.67 (d, J = 8.3 Hz, 24H, H_g), 7.61 (d, J = 7.9 Hz, 24H, H_p), 7.50 – 7.29 (m, 84H, $H_{a,b,h,k,m,n,o}$), 7.22 – 6.98 (m, 84H, $H_{a,b,h,k,m,n,o}$), 5.63 – 5.42 (m, 48H, H_f), 5.27 – 4.89 (m, 24H, H_i), 3.74 (s, 36H, H_q), 3.73 (s, 36H, H_q), 3.56 – 3.32 (m, 48H, H_j). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 176.7, 173.3, 166.6, 159.4, 153.2, 152.8, 138.9, 137.5, 136.8, 135.4, 133.8, 129.8, 128.4, 127.0, 124.8, 122.6, 122.1, 120.1, 119.2, 112.6, 110.9, 110.7, 55.0, 53.1, 28.0. **LR-ESI-MS**: [charge, calculated mass]: *m/z* = 1687.6 [**4.4**$(NTf_2)_4^{4+}$, 1687.0], 1293.9 [**4.4**$(NTf_2)_3^{5+}$, 1293.7], 1031.6 [**4.4**$(NTf_2)_2^{6+}$, 1031.4], 844.2 [**4.4**$(NTf_2)^{7+}$, 844.1], 703.6 [**4.4**$^{8+}$, 703.6]. **HR-ESI-MS**: *m/z* calculated for **4.4**$(NTf_2)_3^{5+}$ = 1293.7062, observed = 1293.7058.

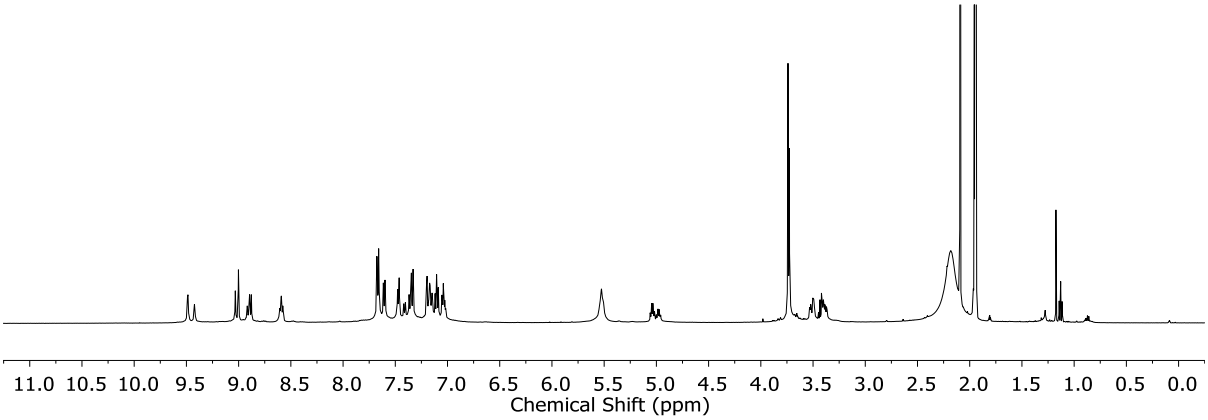


Figure 4.35| ^{1}H NMR (400 MHz, 298K, CD_3CN) of **4.4.**

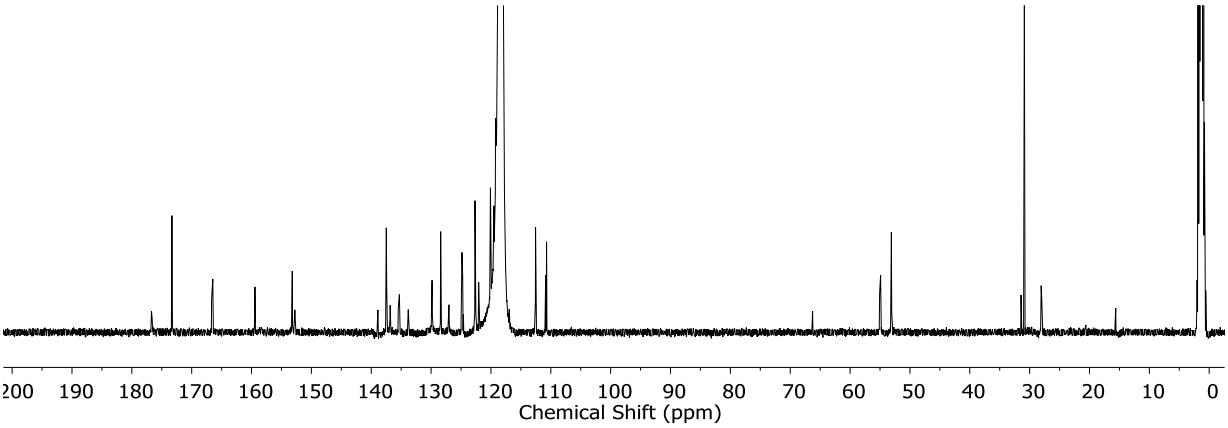


Figure 4.36| ^{13}C NMR (126 MHz, 298K, CD_3CN) of **4.4.**

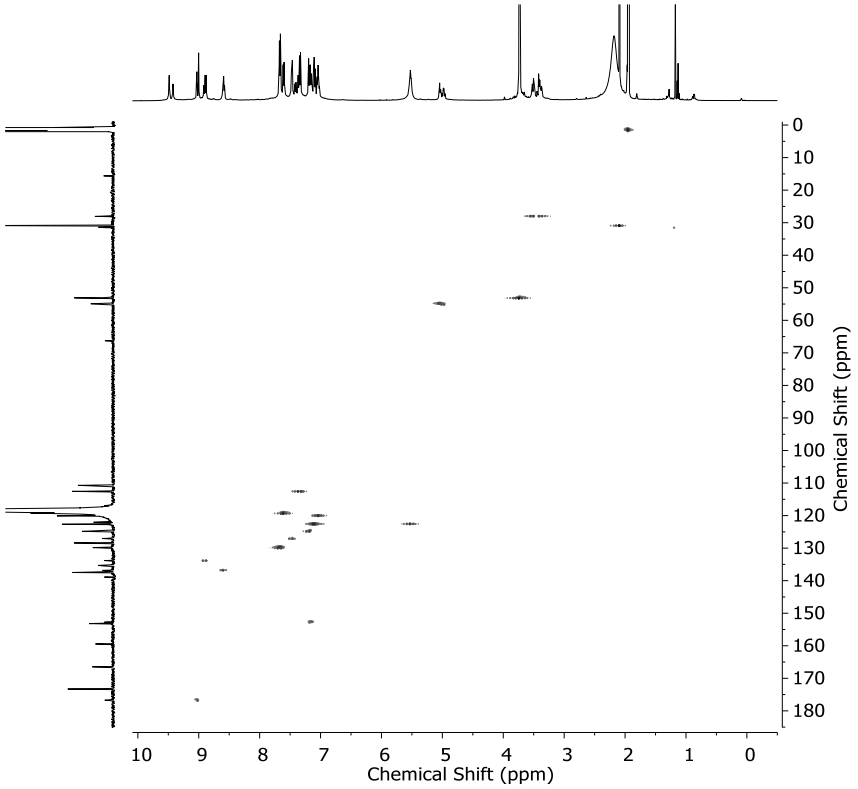
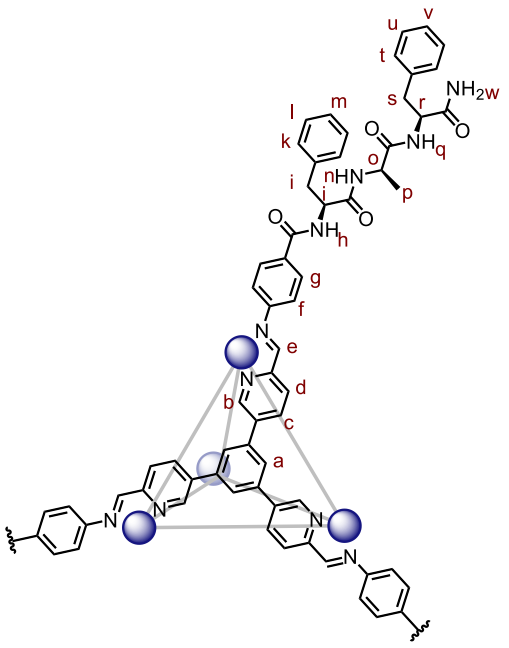


Figure 4.37 | ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, CD_3CN) of **4.4**.

Cage 4.5

The spectra were identical to **4.4.**

Cage 4.6



^{1}H NMR (400 MHz, 298K, CD_3CN) δ 8.92 (s, 12H, H_e), 8.87 (s, 12H, H_e), 8.77 (d, *J* = 8.3 Hz, 12H, H_d), 8.72 (d, *J* = 8.3 Hz, 12H, H_d), 8.48 – 8.41 (m, 24H, H_c), 8.29 (s, 24H, H_h), 7.73 (d, *J* = 8.1 Hz, 24H, H_g), 7.68 (bd, 24H, H_g), 7.49 – 6.94 (m, 240H, $H_{a,b,k,l,m,n,q,t,u,v}$), 6.69 (bs, 48H, H_w), 5.50 (bd, 48H, H_f), 4.77 – 4.66 (m, 24H, H_i), 4.52 (bs, 24H, H_r), 4.25 – 4.10 (m, 24H, H_o), 3.29 – 2.94 (m, 96H, $H_{j,s}$), 1.06 (d, *J* = 6.9 Hz, 36H, H_p), 1.01 (d, *J* = 6.9 Hz, 36H, H_p). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 175.3, 174.3, 172.5, 172.2, 166.2, 157.3, 153.6, 152.4, 137.7, 137.2, 135.5, 135.0, 134.2, 131.9,

129.4, 129.3, 128.4, 128.4, 126.8, 126.6, 123.7, 121.5, 121.1, 55.9, 54.6, 49.6, 37.0, 36.9, 16.5.**^{19}F NMR** (376 MHz, 298K, CD_3CN) δ –79.52, –79.55, –80.29. **LR-ESI-MS** [charge, calculated mass]: *m/z* = 1654.8 [**4.6**(TFA)$(NTf_2)_2^{5+}$, 1654.4], 1621.4 [**4.6**$(TFA)_2(NTf_2)^{5+}$, 1621.0], 1587.8 [**4.6**$(TFA)_3^{5+}$, 1587.6], 1332.0 [**4.6**$(TFA)(NTf_2)^{6+}$, 1332.0], 1304.1 [**4.6**$(TFA)_2^{6+}$, 1304.18], 1101.9 [**4.6**$(TFA)^{7+}$, 1101.7], 949.4 [**4.6**$^{8+}$, 949.9]. **HR-ESI-MS**: *m/z* calculated for **4.6**$(TFA)(NTf_2)_2^{5+}$= 1654.5494, observed = 1654.5491.

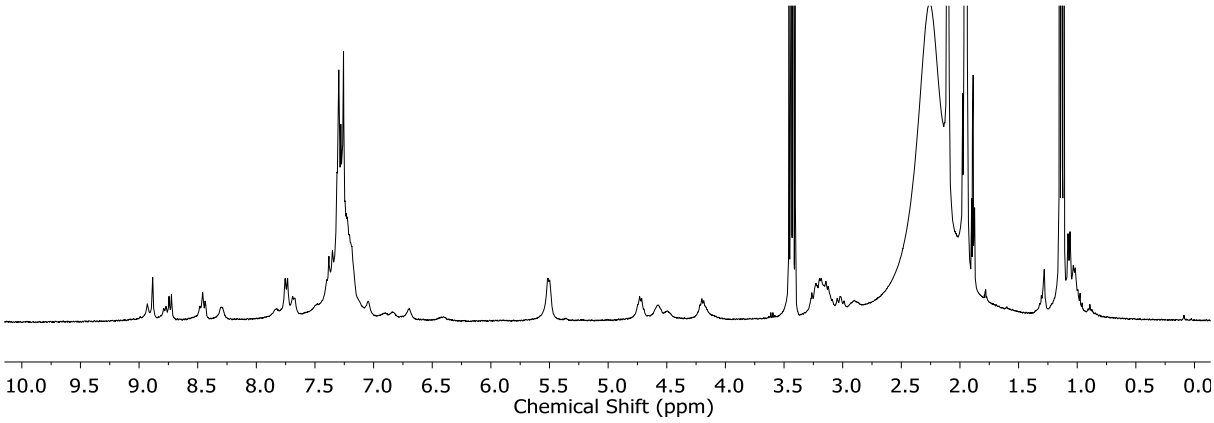


Figure 4.38| ^{1}H NMR (400 MHz, 298K, CD_3CN) of **4.6.**

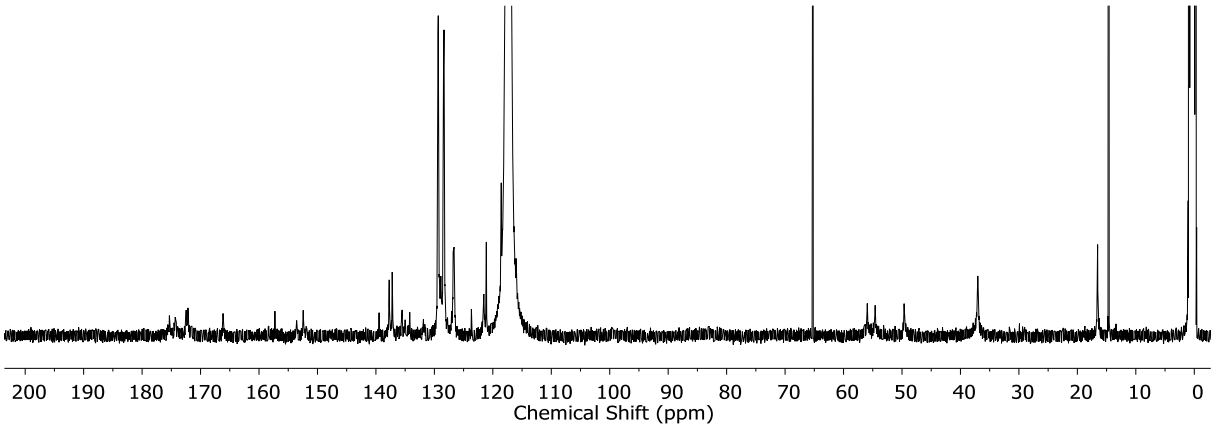


Figure 4.39| ^{13}C NMR (126 MHz, 298K, CD_3CN) of **4.6.** Low concentrations were required to avoid aggregation and subsequent gelation of the compound, resulting in low quality of the ^{13}C spectrum.

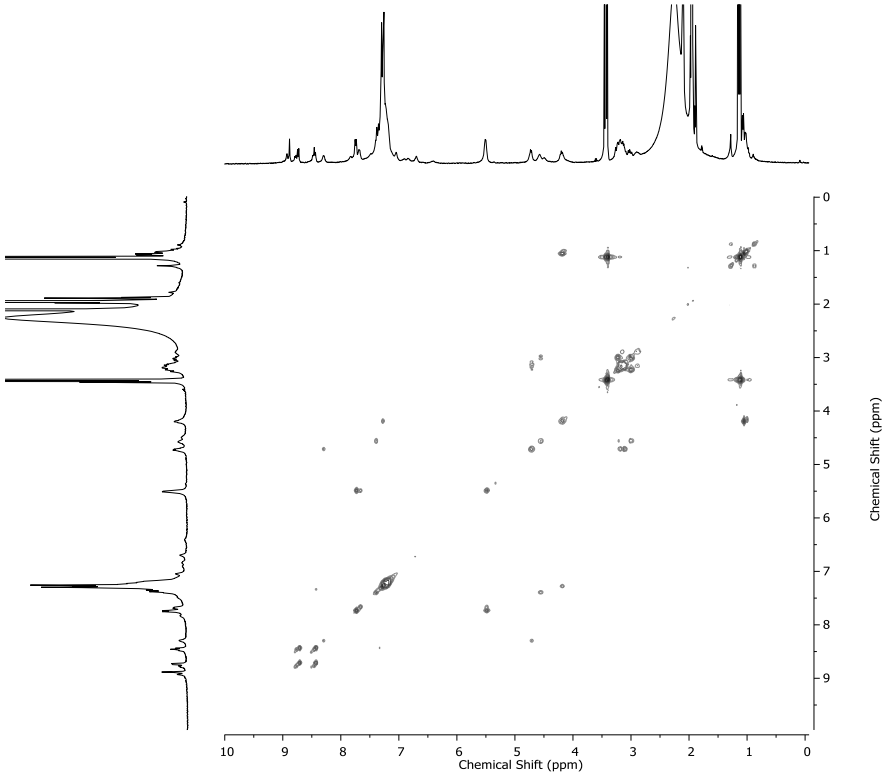


Figure 4.40| ^{1}H-^{1}H COSY spectrum (500 MHz, 298 K, CD_3CN) of **4.6**

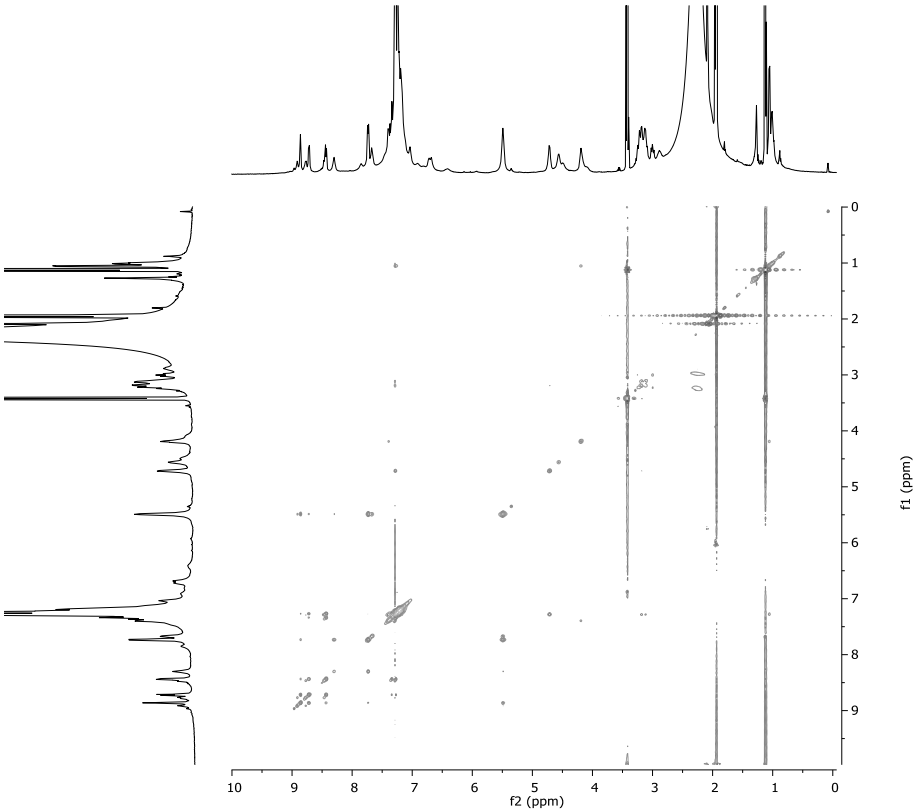


Figure 4.41| ^{1}H-^{1}H NOESY spectrum (500 MHz, 298 K, CD_3CN) of **4.6**

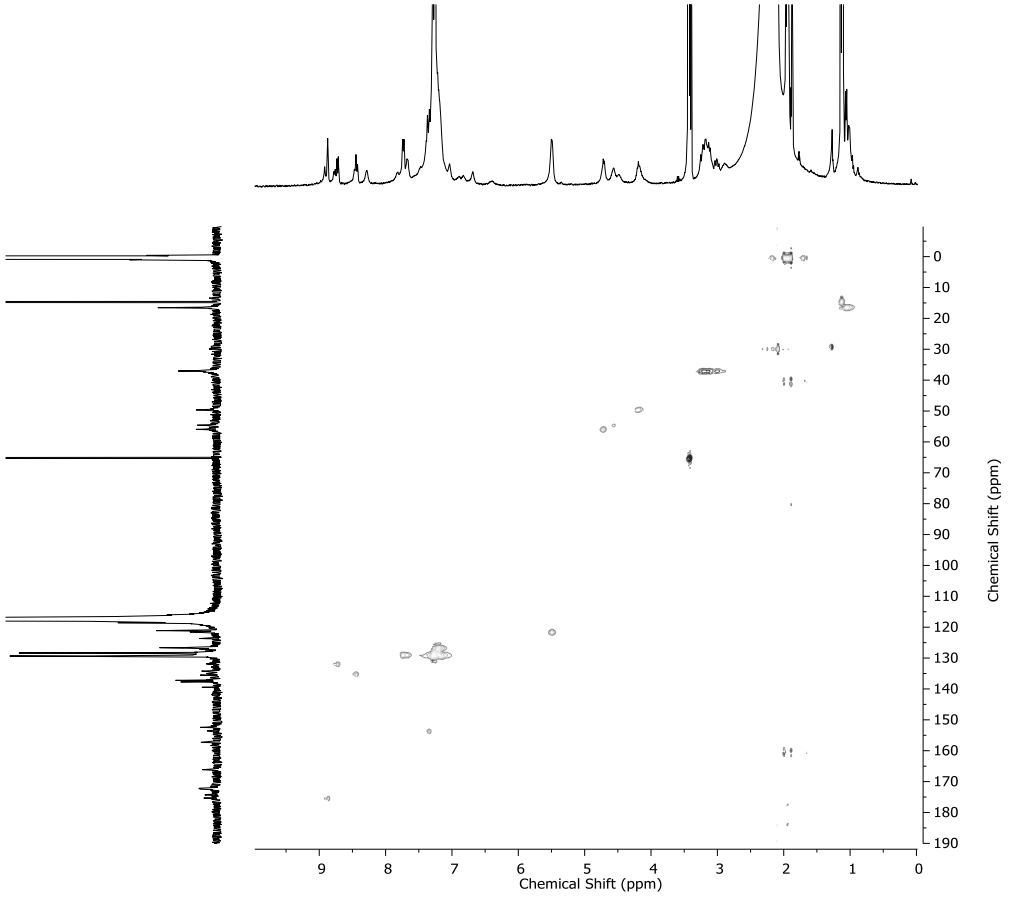


Figure 4.42 | ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, CD_3CN) of **4.6**

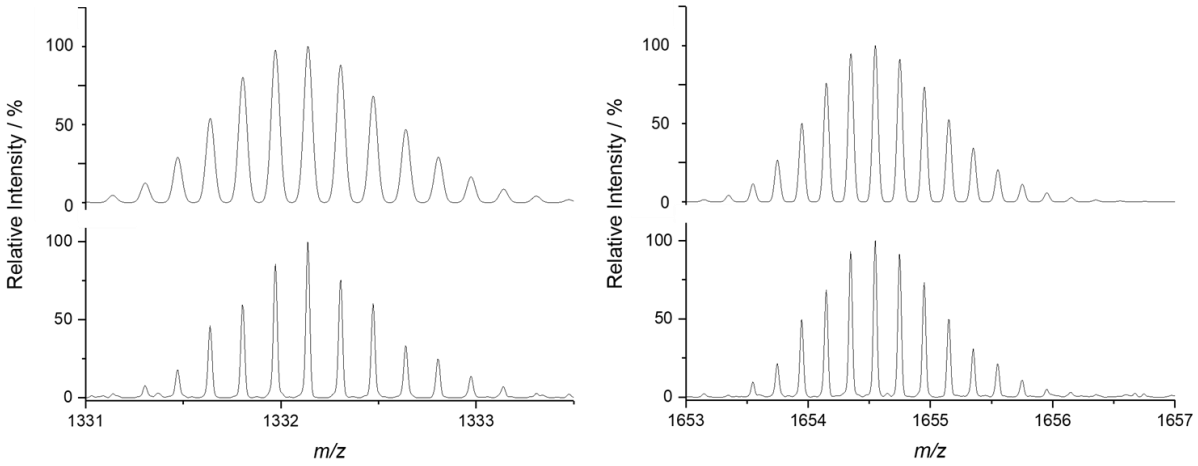


Figure 4.43 | HR ESI-MS mass spectrum of **4.6** showing the observed z = +6 charge, (bottom left) compared to the theoretical isotope pattern (top left) and the observed z = +5 charge, (bottom right) compared to the theoretical isotope pattern (top right).

4.8 References

[1] N. Kitadai, S. Maruyama, *Geosci. Front.* **2018**, *9*, 1117-1153.

[2] B. E. I. Ramakers, J. C. M. van Hest, D. W. P. M. Löwik, *Chem. Soc. Rev.* **2014**, *43*, 2743-2756.

[3] D. Mandal, A. Nasrolahi Shirazi, K. Parang, *Org. Biomol. Chem* **2014**, *12*, 3544-3561.

[4] N. Habibi, N. Kamaly, A. Memic, H. Shafiee, *Nano Today* **2016**, *11*, 41-60.

[5] A. V. Guzzo, *Biophys. J.* **1965**, *5*, 809-822.

[6] M. Yamagami, T. Sawada, M. Fujita, *J. Am. Chem. Soc.* **2018**, *140*, 8644-8647.

[7] K. Suzuki, M. Kawano, S. Sato, M. Fujita, *J. Am. Chem. Soc.* **2007**, *129*, 10652-10653.

[8] H.-Y. Lee, J. W. Kampf, K. S. Park, E. N. G. Marsh, *Cryst. Growth Des.* **2008**, *8*, 296-303.

[9] X. Li, J. Fei, Y. Xu, D. Li, T. Yuan, G. Li, C. Wang, J. Li, *Angew. Chem. Int. Ed.* **2018**, *57*, 1903-1907.

[10] Z. Shi, P. Peng, D. Strohecker, Y. Liao, *J. Am. Chem. Soc.* **2011**, *133*, 14699-14703.

[11] C. Maity, W. E. Hendriksen, J. H. van Esch, R. Eelkema, *Angew. Chem. Int. Ed.* **2015**, *54*, 998-1001.

[12] N. Abeyrathna, Y. Liao, *J. Phys. Org. Chem.* **2017**, *30*, e3664.

[13] J. Guo, H.-Y. Zhang, Y. Zhou, Y. Liu, *Chem. Commun.* **2017**, *53*, 6089-6092.

[14] Q. Shi, C.-F. Chen, *Org. Lett.* **2017**, *19*, 3175-3178.

[15] S. M. Jansze, G. Cecot, K. Severin, *Chem. Sci.* **2018**, *9*, 4253-4257.

[16] T. Zhang, L. Sheng, J. Liu, L. Ju, J. Li, Z. Du, W. Zhang, M. Li, S. X.-A. Zhang, *Adv. Funct. Mater.* **2018**, *28*, 1705532.

[17] N. Zidar, H. Macut, T. Tomašič, M. Brvar, S. Montalvão, P. Tammela, T. Solmajer, L. Peterlin Mašič, J. Ilaš, D. Kikelj, *J. Med. Chem.* **2015**, *58*, 6179-6194.

[18] Y. Qian, A. Vogt, A. Vasudevan, S. d. M. Sebti, A. D. Hamilton, *Biorg. Med. Chem.* **1998**, *6*, 293-299.

[19] F. Stewart, *Aust. J. Chem.* **1983**, *36*, 1629-1638.

Chapter 5

N-Acetylgalactosamine bioconjugated $Fe^{II}_4L_4$ tetrahedral cage to target hepatocytes

5.1 Introduction

Current work in drug-delivery focuses on improving treatments by increasing both the drug uptake in cells and the retention time of the drug in the body. Achieving targeted delivery of drug payload represents the next major step towards new medical treatments with higher potency and overall lower toxicity towards healthy cells. To do so, the main strategy relies on appending cell-targeting or cell-translocating groups, like cyclic-RGD (cyclic peptides containing the sequence Arg-Gly-Asp) for example, on the drug delivery vehicle to increase its accumulation in the targeted organs or cells. So far, only a few examples of bioconjugated coordination cages are reported, despite the advantages they could present in applications such as drug delivery (see Chapter 1, sections 1.2.4 and 1.3.3). Amongst those, only one recent example was used for drug delivery, and demonstrated only a small enhancement in activity of the encapsulated drug.[1]

The asialoglycoprotein receptor (ASGPR) is a lectin (carbohydrate binding) transmembrane protein expressed exclusively at the surface of hepatocytes and is involved in the regulation of glycoproteins in the blood. The receptor binds specifically to molecules that contain terminal galactose or *N*-acetylgalactosamine residues, leading to their internalization and subsequent removal from blood circulation. Once bound to ASGPR, very large molecules can be internalized in the cells *via* endocytosis. Furthermore, good selectivity can be achieved by targeting this receptor, as high numbers are found on the surface of liver cells (500,000 ASGPR/hepatocyte) but it is minimally expressed in the rest of the body.[2]

The spatial geometry of the ligands is crucial in the binding process to ASGPR, with naturally occurring highly branched carbohydrates polymers or glycoprotein and glycopeptides being particularly successful.[3] Over the years, synthetic multivalent ligands with differing branching, hydrophobicity, spatial arrangement and terminating sugars have been screened for binding to the receptor.[4] The most successful binders were shown to result from a symmetric triangular arrangement of *N*-acetylgalactosamine (GalNAc) held 19 Å to 32 Å apart (Figure 5.1 a and b).

Scrutiny of *fac*-complexes formed by an octahedral metal and three bidentate pyridyl-imine ligands showed that substituents at the 5-position on the pyridine ring (Figure 5.1 c, green spheres) are placed in a triangular geometry. It was hypothesized that using a picolinaldehyde biotagged with GalNAc as a building block in the self-assembly of tetrahedra displaying *fac*-vertices would result in coordination cages able to bind selectively to ASGPR.

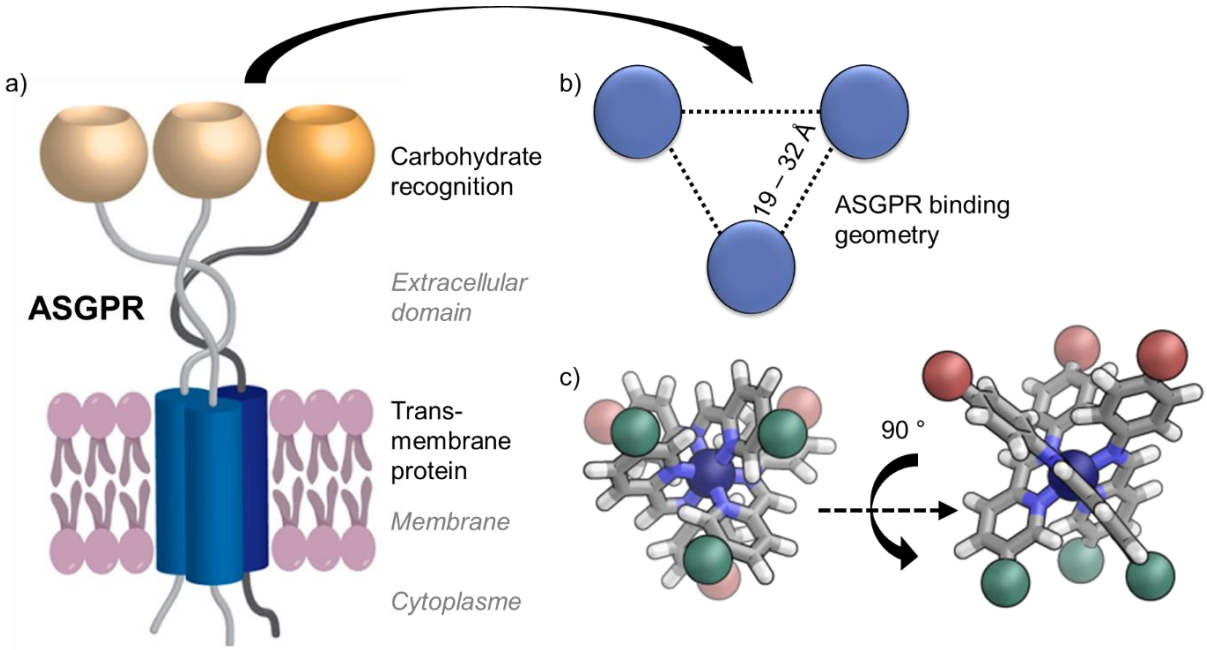


Figure 5.1 | a) Cartoon representation of ASGPR receptor depicting the transmembrane domain (blue) to which three carbohydrate recognition sites (gold) are connected and b) the resulting geometry of the binding sites of ASGPR, with the distance between site varying from 19 to 32 Å. c) Cartoon representation of the crystal structure of a *fac* metal complex from an octahedral metal and three bidentate ligands, showing the triangular arrangement of the groups at the 5-position of the pyridine. The scale is different between a) and c). Figure a) adapted with permission from reference 4.

In this chapter, a new strategy to attach targeting unit on coordination cages was explored, using the versatility of sub-components self-assembly. A picolinaldehyde bioconjugated to a GalNAc was synthesized and incorporated in water-soluble tetrahedral cages. Similarly, a fluorescent dye was conjugated to a picolinaldehyde and incorporated in the coordination cage tagged with GalNAcs. The stability of the cages in water and buffer was tested, showing stability in PBS down to micromolar concentrations. The binding of the coordination cages to hepatocytes *via* the ASGPR will be tested next.

5.2 Synthesis and design of subcomponents

5.2.1 Synthesis of the GalNAc biotag

As mentioned above, previous design of ligands for ASGPR binding identified molecules with 19 – 32 Å spacing between the sugars as being most efficient.[4] For a triangular arrangement, this spacing can be obtained by linkers ranging from 12 to 20 Å (measured between the C-1 of the sugar and the connection point of the three arms). Based on these considerations, target **5.J**

was designed, comprising of a GalNAc for binding to the ASGPR, a 19 Å hydrophobic linker and a picolinaldehyde moiety for cage assembly (Figure 5.2).

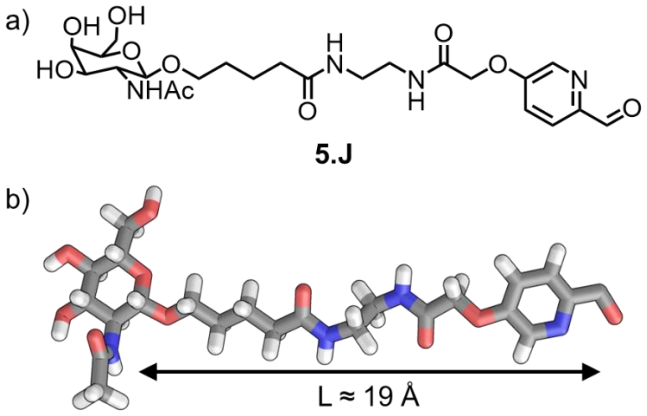
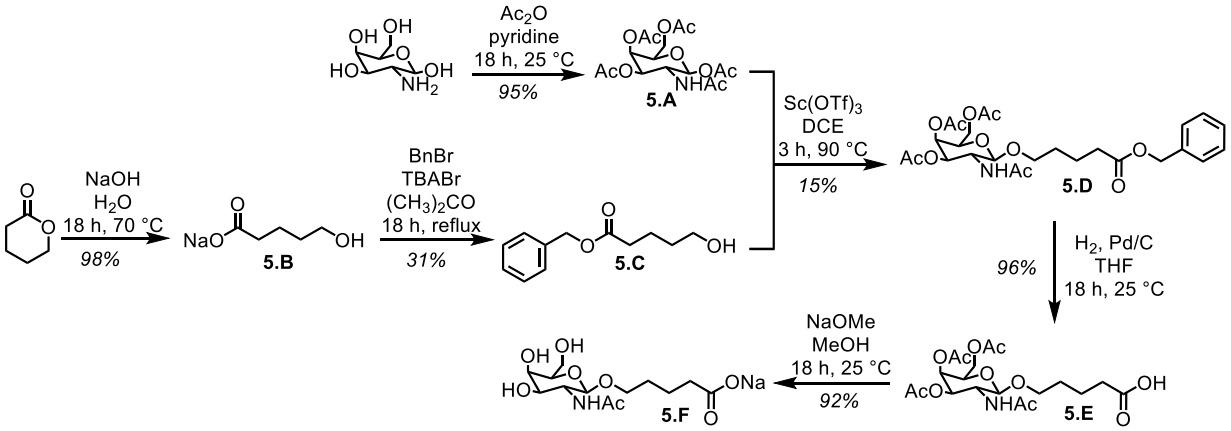


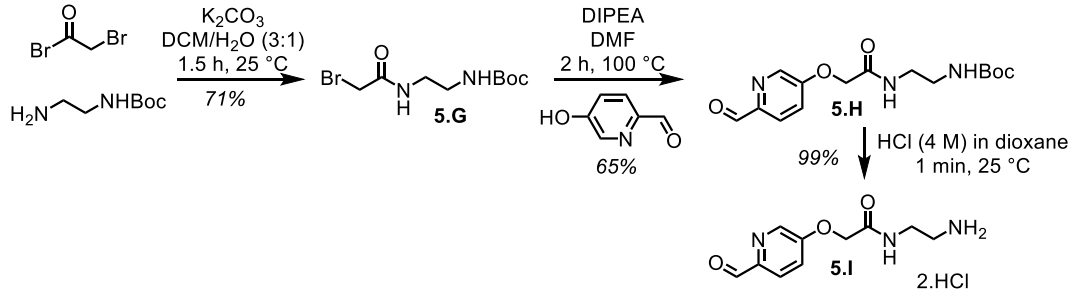
Figure 5.2 | a) Synthetic target **5.J** and b) MM3-optimised model of **5.J**

Compound **5.F** was prepared by a synthesis adapted from an existing patent (Scheme 5.1).[5] D-galactosamine pentaacetate **5.A** was obtained by reaction of D-galactosamine with acetic anhydride in anhydrous pyridine. The hydrolysis of δ-valerolactone in the presence of sodium hydroxide yielded compound **5.B**. The reaction of benzyl bromide with **5.B** in acetone in the presence of a phase-transfer catalyst (TBABr) yielded multiple products, which could be isolated by silica gel chromatography. The various products were identified by NMR and LCMS to be the result of the polymerisation of **5.B** and **5.C** (up to 4-mer). Despite attempts to alter the reaction conditions, no conditions could be found, which decreased the amount of polymers, leading to low yield in this step. The reaction of **5.C** and **5.A** in the presence of $Sc(OTf)_3$ proved to have a high anomeric selectivity, forming 95% of the β-D-galactopyranoside. However, polymerisation of the benzyl 5-hydroxypentanoate linker **5.C** was observed once more, leading to a library of sugars functionalised with side chains containing up to six linkers. The desired compound **5.D** was isolated by preparative HPLC in low yield. Subsequent deprotection of the carboxylic acid followed by deacetylation of the sugar gave compound **5.F** in near quantitative yield.

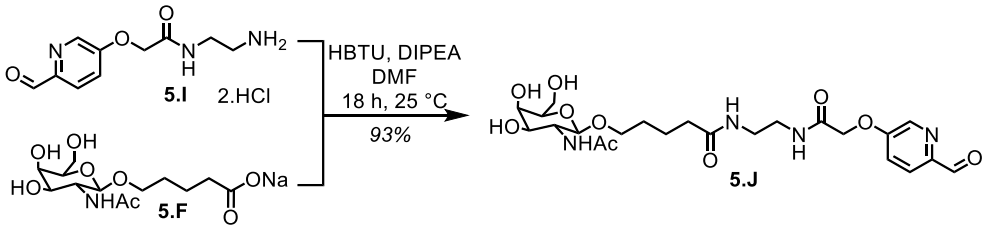
The reaction of 2-bromoacetic acid with *tert*-butyl (2-aminoethyl)carbamate in basic media yielded compound **5.G**. Compound **5.H** was obtained by nucleophilic substitution of the commercial 5-hydroxypicolinaldehyde with **5.G**. Removal of the Boc protecting group with HCl (4 M) in dioxane yielded the target compound **5.I** (Scheme 5.1). Finally, amide coupling between **5.I** and **5.F** in the presence of HBTU and DIPEA gave the desired target **5.J** in near quantitative yield (Scheme 5.3).



Scheme 5.1 | Synthetic scheme for **5.F**.



Scheme 5.2 | Synthetic scheme for **5.I**.



Scheme 5.3 | Synthetic scheme for the final target compound **5.J**.

5.2.2 Synthesis of the fluorescent tag

Most supramolecular architectures obtained in the Nitschke group have no fluorescent properties, and those which do could not be transferred into water successfully. Consequently, a novel fluorescent tag was designed for appendage to the cage. The autofluorescence of human hepatocytes (HepG2) in the green region of the spectrum[6] prevents the use of common fluorophores such as fluorescein, Lucifer yellow or BODIPY.[7] The use of the popular commercial dye Alexa Fluor™ 647[8] was limited due to its high price, challenging synthesis

and high negative charge of three which could reduce the cage's solubility by lowering the overall charge of the structure. The less common cyanine dyes with five and seven carbon atoms between the indole moieties (Cy5 and Cy7) have emission wavelengths of 670 nm and 767 nm respectively and a positive charge which can contribute to the solubility of the cages. However, we expect these compounds to be poorly soluble in water and therefore unsuitable for our system.

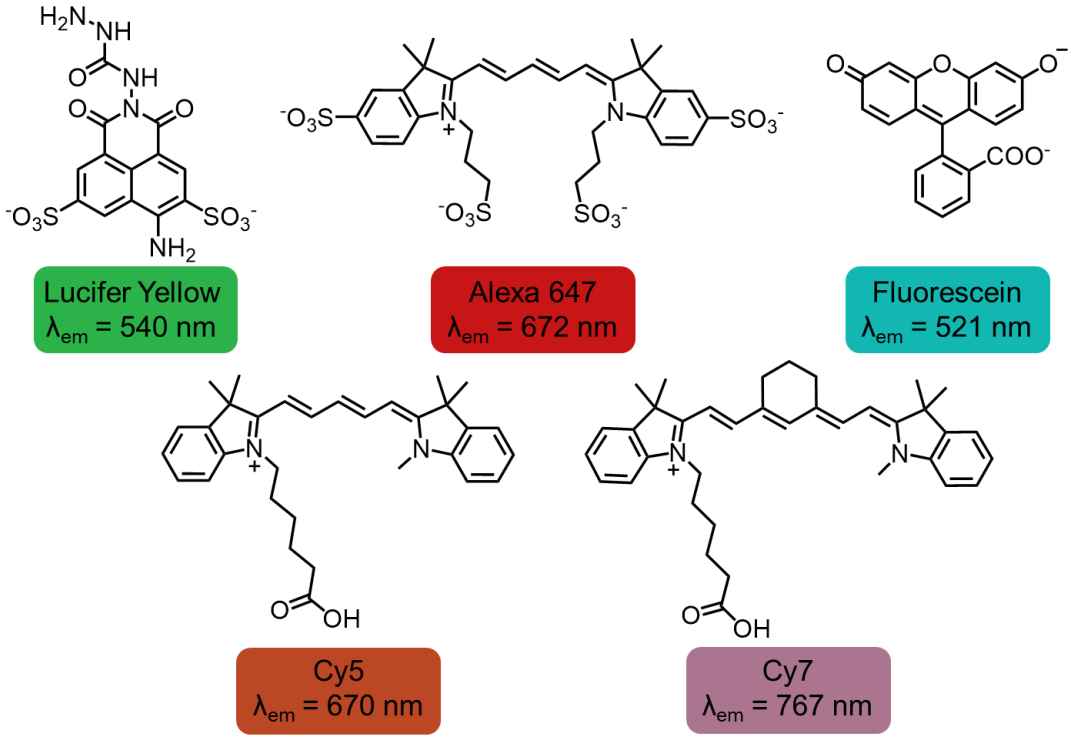
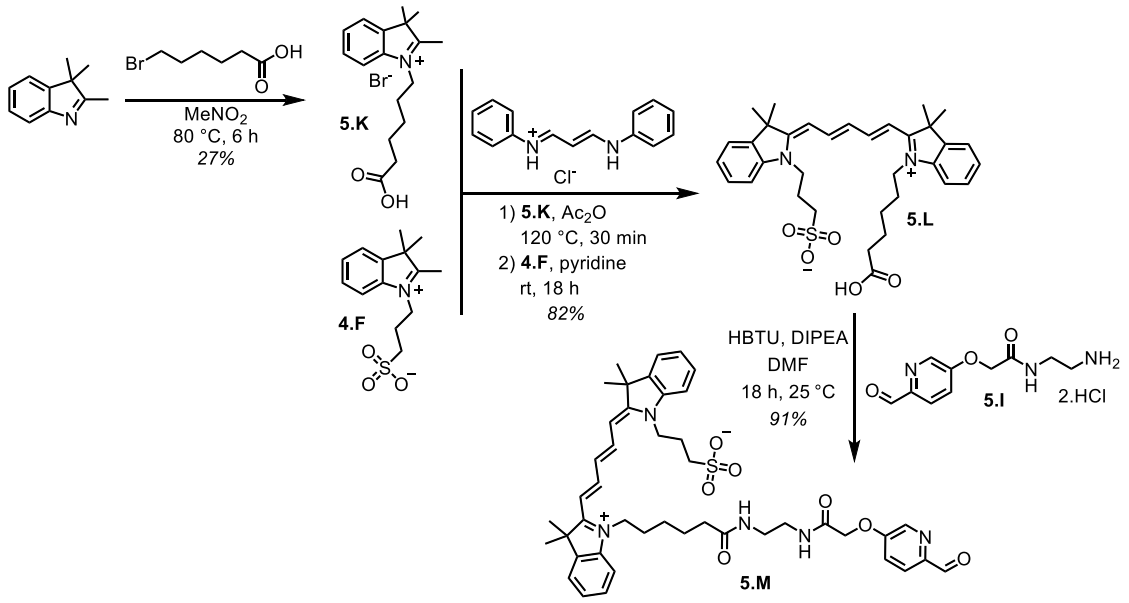


Figure 5.3 | Structure of fluorescent dyes and their emission wavelength in water.

Recently, an alternative fluorescent dye (**5.L**) based on the cyanine structure was developed.[9] This new compound has a sulfonated chain instead of the methyl on the indole, which increases its water solubility. Dye **5.L** was synthesised from an adapted literature procedure (Scheme 5.4).[10] The synthetic target **5.M** was designed based on the same principle as **5.J**, comprising of both a picolinaldehyde for attachment to the cage and the fluorescent dye **5.L**.

The fluorescence spectrum of **5.M** confirmed that the addition of a pyridine moiety did not quench the fluorescence. The compound formed appeared to be suitable for use in cell studies as both the maximum emission wavelength (671 nm) and maximum excitation (659 nm) in water were out of the green channel (Figure 5.4 a). The additional band at 580 nm in the excitation spectra was inferred to result from aggregation due to the low solubility of the compound in water.



Scheme 5.4 | a) Synthetic scheme for the alternative fluorescent dye **5.L** and synthetic target **5.M**.

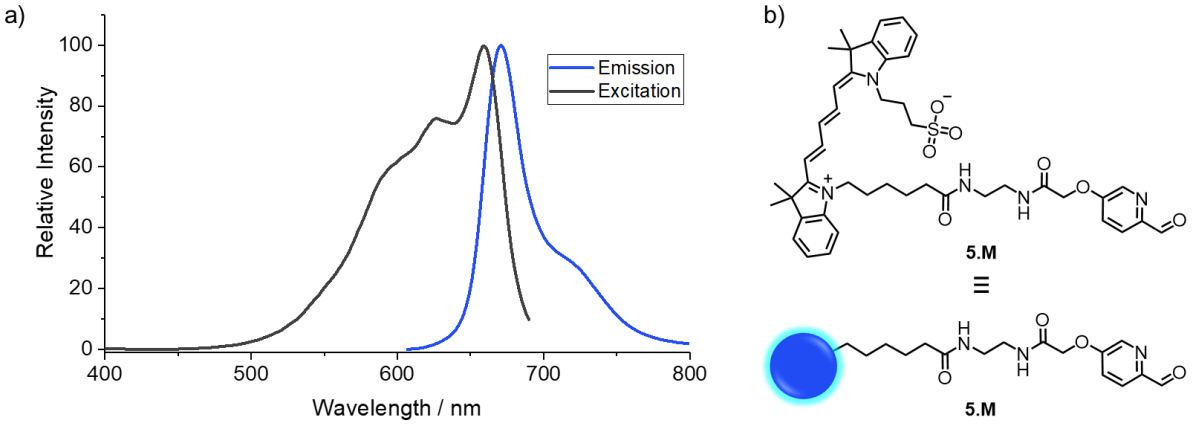


Figure 5.4 | a) Fluorescence spectra of **5.M** (excitation in black and emission in blue). b) Structure and schematic representation of **5.M**.

5.3 Formation of supramolecular coordination cages and stability in biological media

To be suitable for cell studies, the coordination cages employed need to be both soluble and stable in water. In addition, metals such as Co^{II} or Cd^{II} could not be used due to their toxicity and Fe^{II} was thus favoured. Three trianilines known to readily form cages in water when assembled with picolinaldehyde and $FeSO_4$ were tested for self-assembly with **5.J**. Neither **3.B** nor the pararosaniline base assembled into a discrete tetrahedral cage when combined with **5.J**, with the unreacted subcomponents being the main species observed in the ^1H NMR spectra of the crude reaction mixture.

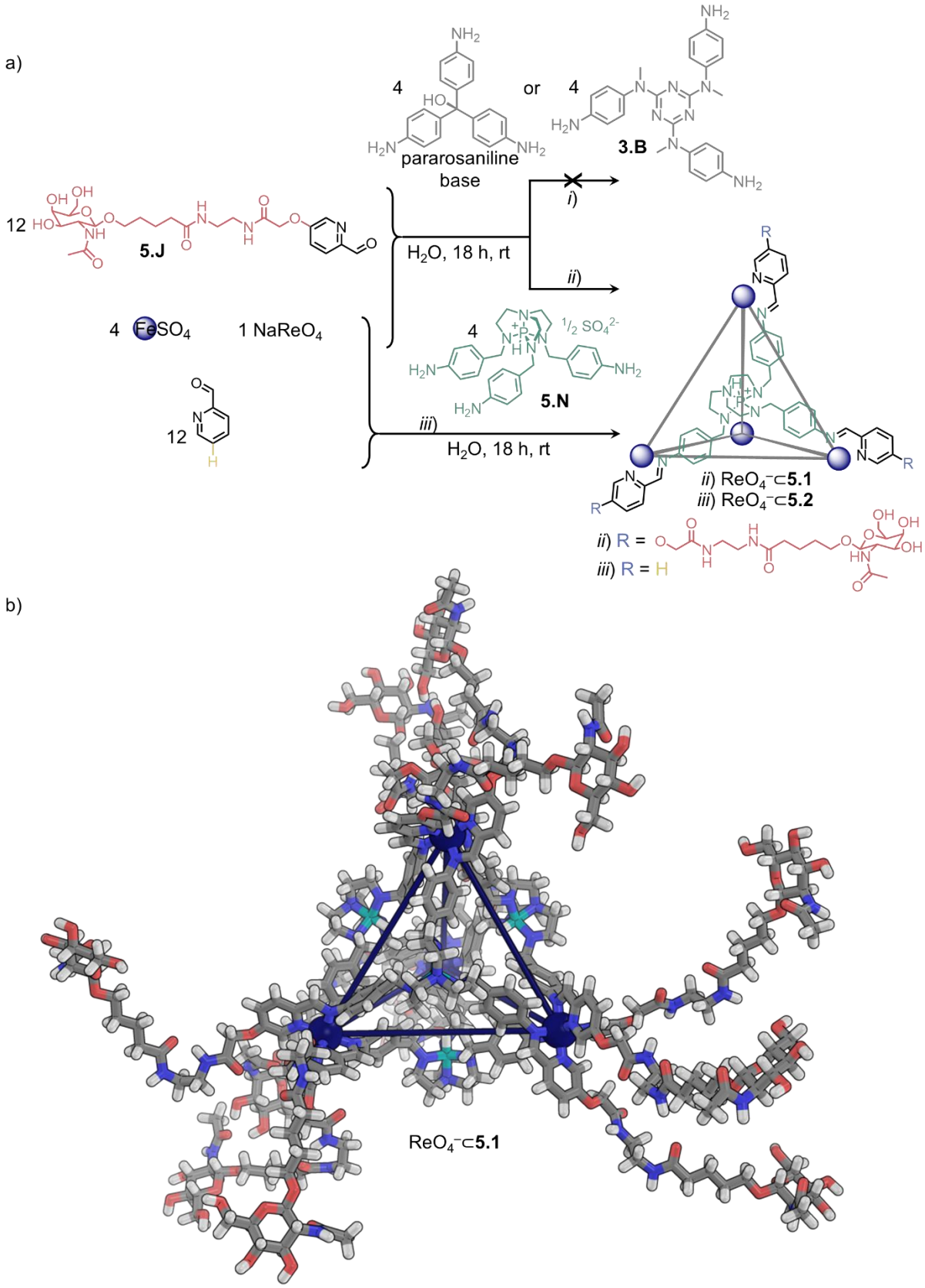


Figure 5.5 | a) Subcomponent self-assembly from i) **5.J** (12 eq.), pararosaniline base (4 eq.) or **3.B** (4 eq.), $FeSO_4$ (4 eq.) and $NaReO_4$ (1 eq.) in D_2O giving mainly subcomponents, ii) **5.J** (12 eq.), **5.N** (4 eq.), $FeSO_4$ (4 eq.) and $NaReO_4$ (1 eq.) in D_2O giving $ReO_4^- \subset$ **5.1** and iii) picolinaldehyde (12 eq.), **5.N** (4 eq.), $FeSO_4$ (4 eq.) and $NaReO_4$ (1 eq.) giving $ReO_4^- \subset$ **5.2** b) MM3-optimised model of $ReO_4^- \subset$ **5.1**. The encapsulated ReO_4^- was omitted for clarity.

When **5.J** (12 eq.) was combined with **5.N** (4 eq.), $FeSO_4$ (4 eq.) and $NaReO_4$ (1 eq.) in D_2O at room temperature for 18 h, $ReO_4^-\subset$**5.1** was formed as a discrete species in solution (Figure 5.5 a). The ^1H NMR was similar to the previously reported $Fe^{II}_4L_4$ tetrahedral cage $ReO_4^-\subset$**5.2**, formed with picolinaldehyde under the same conditions (Figure 5.6).[11] HSQC NMR showed the absence of signals matching with aldehyde protons and the presence of a signal matching with an imine proton, indicative of the condensation of **5.J** and **5.N**. No mass spectra confirming the stoichiometry of cage $ReO_4^-\subset$**5.1** were obtained, a recurring issue with this type of water soluble coordination cages.[12] MM3-optimised model showed that the $Fe^{II}_4L_4$ tetrahedral cage $ReO_4^-\subset$**5.1** could form as expected with no significant strain (Figure 5.5 b).

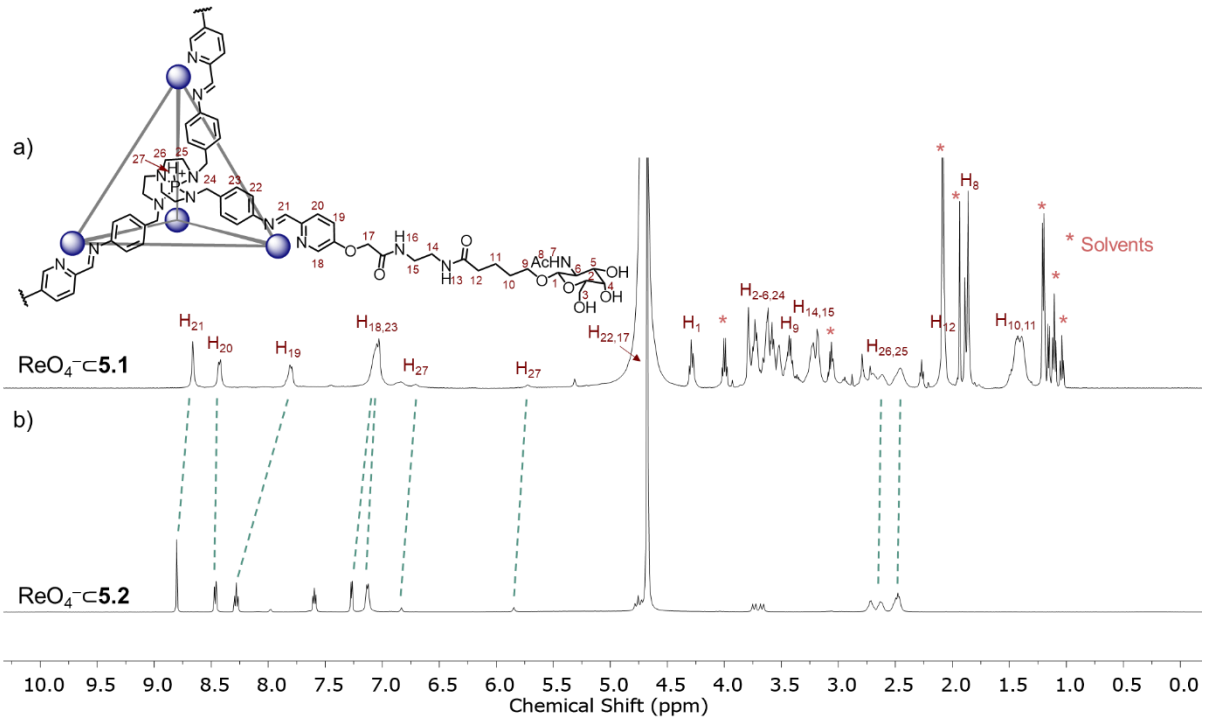


Figure 5.6 | ^1H NMR spectra (500 MHz, 298K, D_2O) of a) $ReO_4^- \subset$ **5.1** with assignment and b) $ReO_4^- \subset$ **5.2**, showing the similarity between the two complexes.

When the self-assembly was carried in the presence of PF_6^- and NTf_2^- as templates instead of ReO_4^-, signals for the encapsulated anions in slow exchange on the NMR time scale were observed in the ^{19}F NMR. Free NTf_2^- was also observed due to its presence in excess and its lower binding affinity than PF_6^-. ^1H NMR revealed the presence of two cage species, attributed to $NTf_2^-\subset$**5.1** and $PF_6^-\subset$**5.1**. Both ^{19}F and ^1H NMR spectra were consistent with those observed for $NTf_2^-\subset$**5.2** and $PF_6^-\subset$**5.2** in the literature.[11] This was indicative of the formation of a cage with a cavity of the appropriate size to encapsulate these anions, likely similar to the $Fe^{II}_4L_4$ tetrahedron **5.2** previously reported.

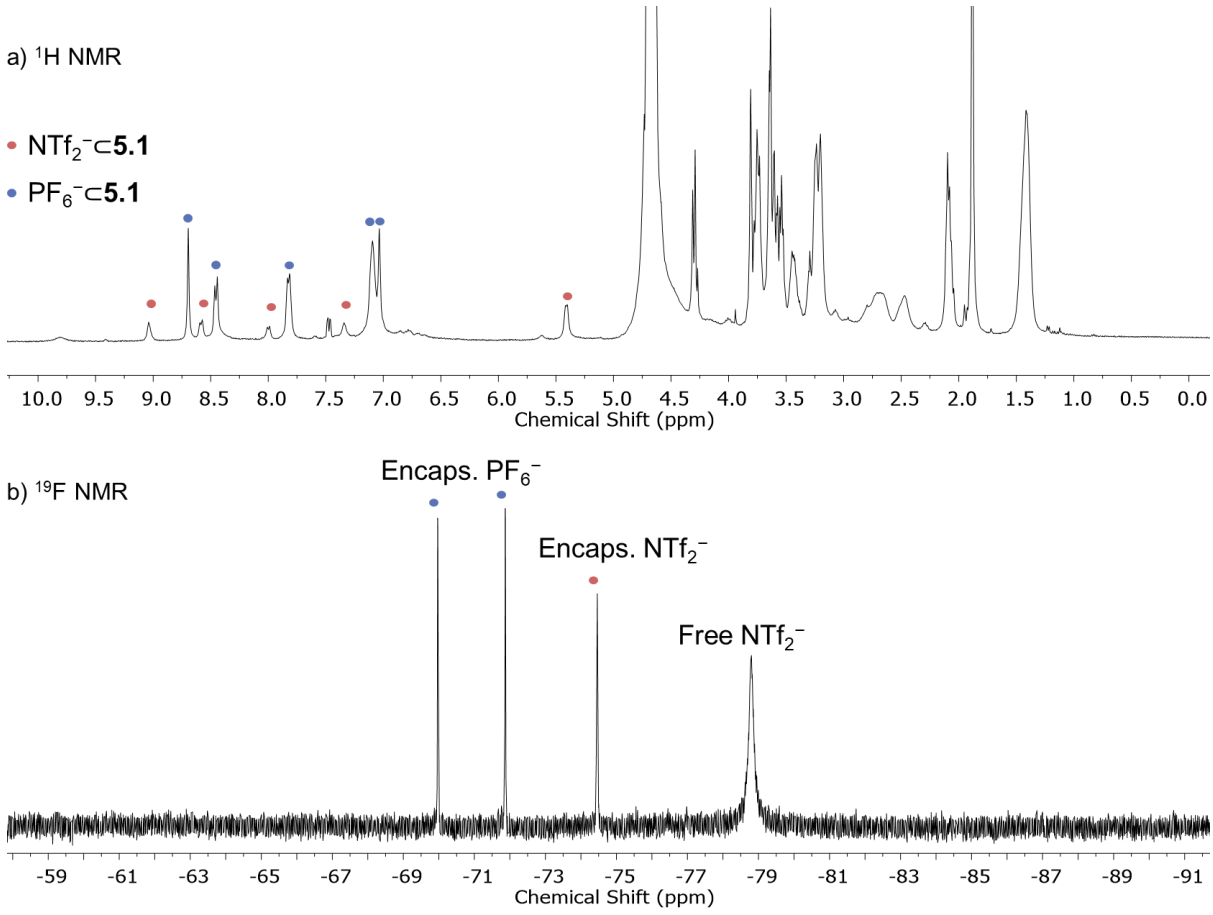


Figure 5.7 | a) ^{1}H NMR spectrum (400 MHz, 298 K, D_2O) of NTf_2^- ⊂ **5.1** (red dots) and PF_6^- ⊂ **5.1** (blue dots) and b) ^{19}F NMR spectrum (376 MHz, 298 K, D_2O) of NTf_2^- ⊂ **5.1** (red dots) and PF_6^- ⊂ **5.1** (blue dots)**,** showing the encapsulated PF_6^- and encapsulated NTf_2^- as well as some free NTf_2^- signals.

Diffusion ordered spectroscopy (DOSY) was further used to probe the formation of a large supramolecular assembly. Proton signals for the pyridine, the aniline, the azaphosphatrane moiety, the GalNAc and the linkers were observed to diffuse at the same speed (Figure 5.8). The diffusion coefficient found for ReO_4^- ⊂ **5.1** was D = 1.10 × 10^{-6} cm^2 s^{-1} from which the radius of the cage was calculated using the Stokes-Einstein equation (Equation 5.1).

Equation 5.1

$$r = \frac{k_B T}{6\pi\eta D} = \frac{1.38\times10^{-23}\times298}{6\pi*8.9\times10^{-4}\times1.10\times10^{-10}} = 2.23\times10^{-9}m$$

Measures on the MM3-optimized model gave an approximate radius of 12 Å for the core of the cage (excluding the GalNAc arms), and an overall radius of 20 – 25 Å including the arms, which is in accordance with the value from the DOSY experiment.

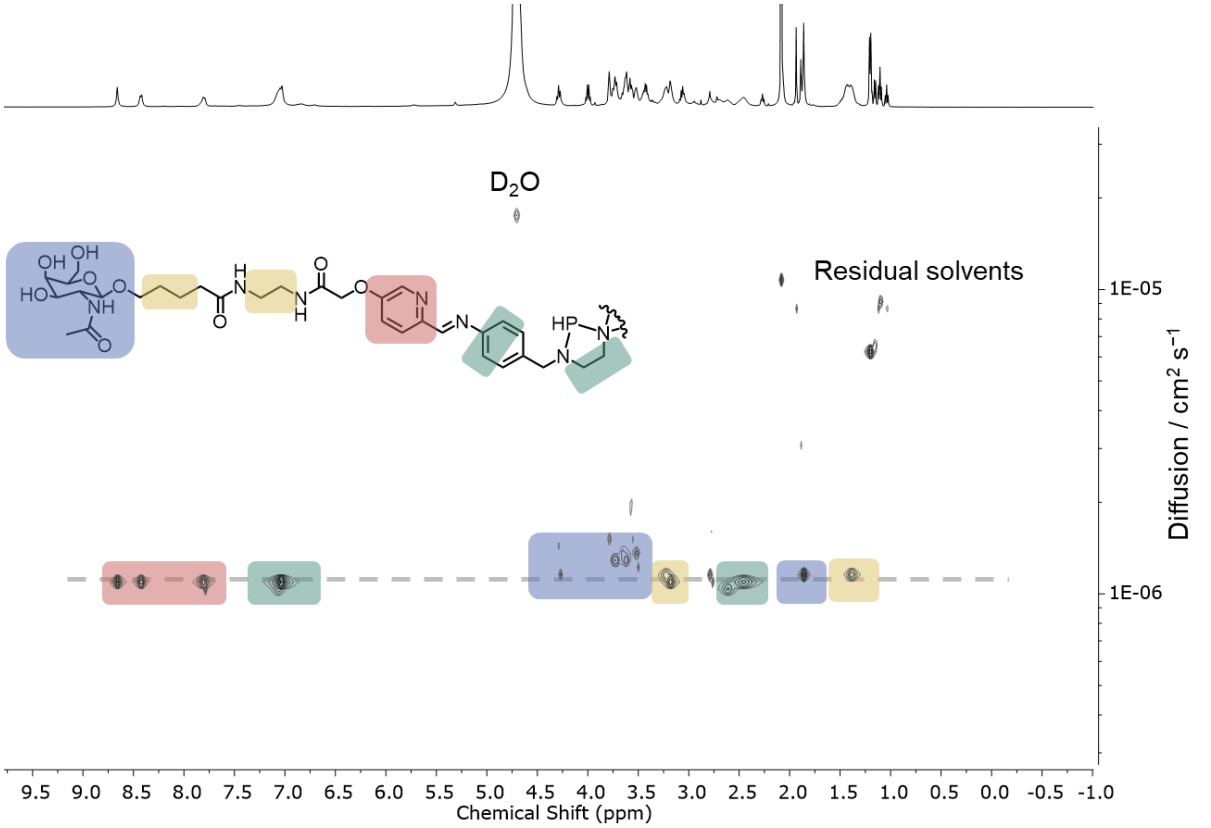


Figure 5.8 | DOSY of ReO_4^-⊂**5.1** with coloured peaks corresponding to the aniline and azaphosphatrane **5.N** (green), pyridine **5.J** (red), GalNAc **5.J** (blue), and linkers **5.J** (yellow). The scatter of the signals was attributed to the overlapping signals and the non-ideal shape of the peaks and baseline.

The formation of ReO_4^-⊂**5.1** was observed to be concentration dependent, with increasing amounts of unreacted subcomponents observed when concentrations of ReO_4^-⊂**5.1** below 1 mM were tested. This effect was explained by the hydrophilicity of subcomponents **5.J** and **5.N** as well as the higher propensity of imine hydrolysis in water. In light of these results, the stability of ReO_4^-⊂**5.1** was tested at low concentrations by UV-vis studies. After formation of cage ReO_4^-⊂**5.1** (2.5 mM), the solution was diluted with water to either 1, 10 or 100 μM and the changes in the MLCT band of cage ReO_4^-⊂**5.1** (maximum absorbance at 546 nm) were monitored over time.

The intensity of the MLCT band decreased over time in all three cases. By normalising the data to unity at t = 0, and under the assumption that once completely decomposed no signal for the MLCT band remained, the relative intensity of the band was assumed to be proportional to the amount of ReO_4^-⊂**5.1** present in solution. After an initial fast decomposition, the decomposition of cage ReO_4^-⊂**5.1** as a function of time was linear. After 14 h, 57% of cage remained for the sample which started at 1 μM, 81% for 10 μM and 90% for 100 μM, showing reasonable stability of cage ReO_4^-⊂**5.1** above 10 μM.

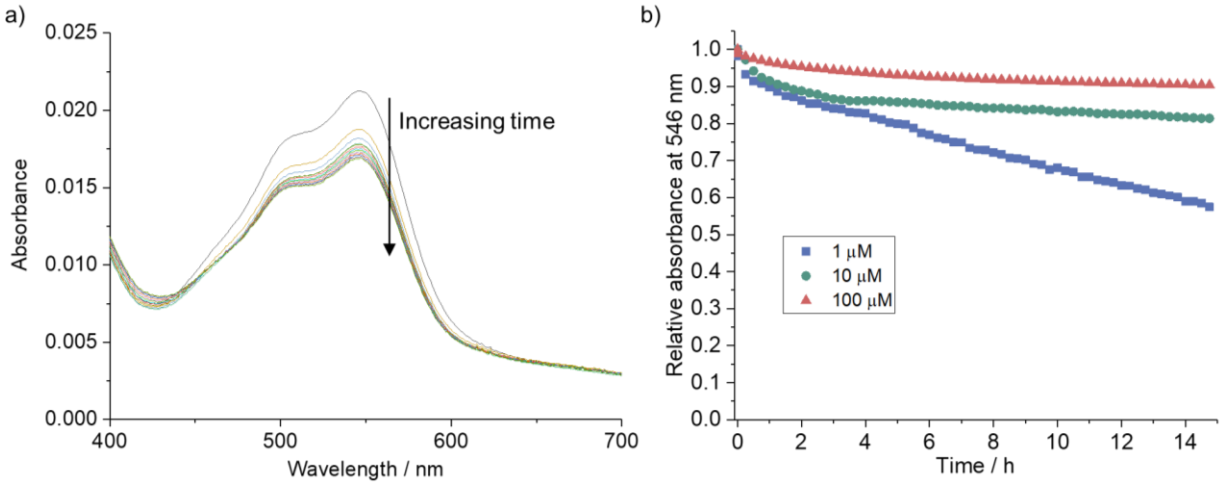


Figure 5.9 | UV-vis spectra in the 400 – 700 nm region corresponding to the MLCT of ReO_4^-⊂**5.1** at 10 uM, showing the changes of the MLCT band over time (one spectrum every hour). b) Plot of the normalised intensity at 546 nm over time for starting concentrations of ReO_4^-⊂**5.1** of 1 μM (blue squares), 10 μM (green circles) and 100 μM (red triangles).

Discrete self-assembled complexes were obtained when **5.J** (10 eq.), **5.M** (2 eq.), **5.N** (4 eq.), $FeSO_4$ (4 eq.) and $NaReO_4$ (1 eq.) were combined in D_2O/CD_3CN (5:1) for 18 h, followed by evaporation of the CD_3CN (Figure 5.10 a). The formation of a library of cages incorporating varying numbers of **5.M** was expected in this case. For ease of notation, this library will be denoted ReO_4^-⊂**5.3**. The ^1H NMR (Figure 5.10 b) and DOSY (Figure 5.11) of the mixture revealed the presence of both GalNAc and dye residues on the coordination cages and they were very similar to the spectra obtained for ReO_4^-⊂**5.1**. The estimated radius of cages ReO_4^-⊂**5.3** from DOSY was 20.8 Å ($D = 1.18 \times 10^{-6}$ cm^2 s^{-1}), which is similar to the value obtained for ReO_4^-⊂**5.1** previously.

The theoretical amount of **5.M** incorporated into cages ReO_4^-⊂**5.3** was estimated by the probability mass function (Equation 5.2).

Equation 5.2

$$P\,(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

With $0 < k < 12$, $p = {}^1/_6$ and $n = 12$. The calculations indicated that more than 87% of all cages formed have fewer than 3 dyes incorporated (Figure 5.12), limiting the probability of self-quenching as well as the chances of having no metal vertices where three GalNAc are attached.

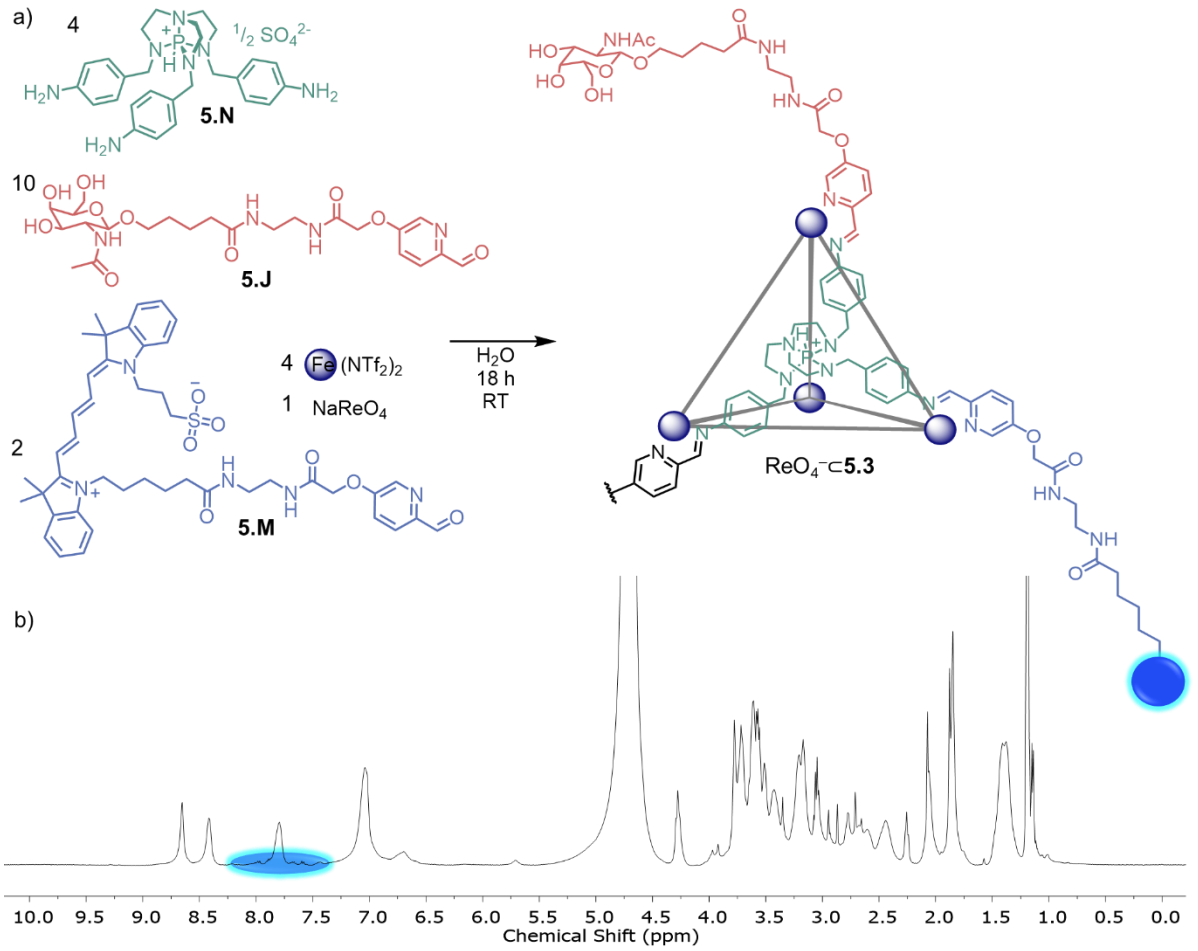


Figure 5.10 | a) Subcomponent self-assembly of ReO_4^-⊂**5.3** from **5.J** (10 eq.), **5.M** (2 eq.), **5.N** (4 eq.), $FeSO_4$ (4 eq.) and $NaReO_4$ (1 eq.) in D_2O/CH_3CN (5:1). b) ^1H NMR (500 MHz, 298 K, D_2O) of ReO_4^-⊂**5.3** with the signals for the dye highlighted in blue.

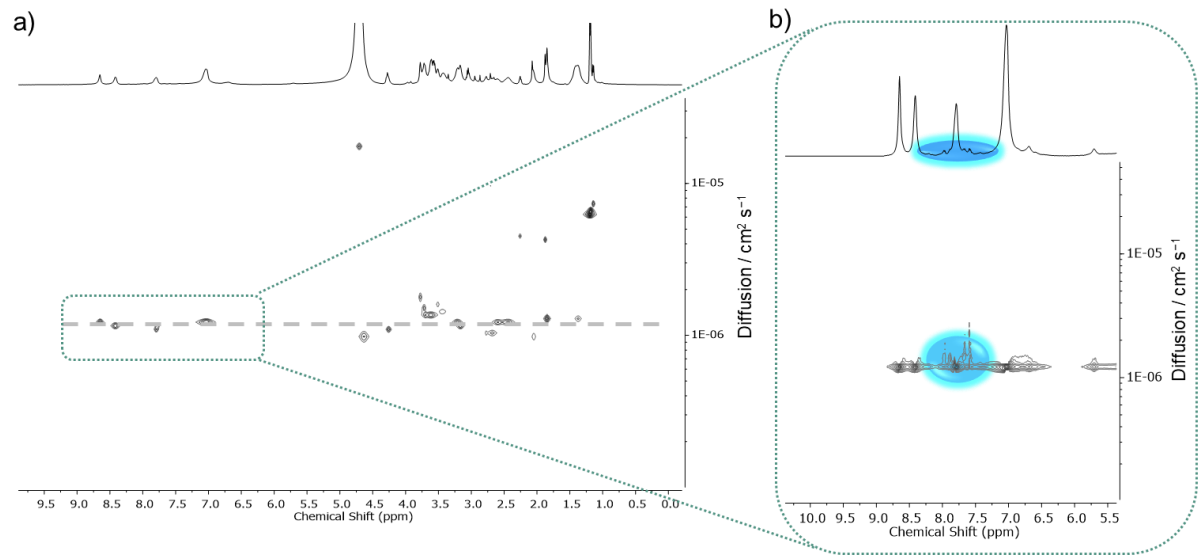


Figure 5.11 | a) DOSY spectrum of ReO_4^-⊂**5.3** and b) expansion of the aromatic region showing the signals for the dye (blue). The scatter of the signals was attributed to the overlapping signals and the non-ideal shape of the peaks and baseline.

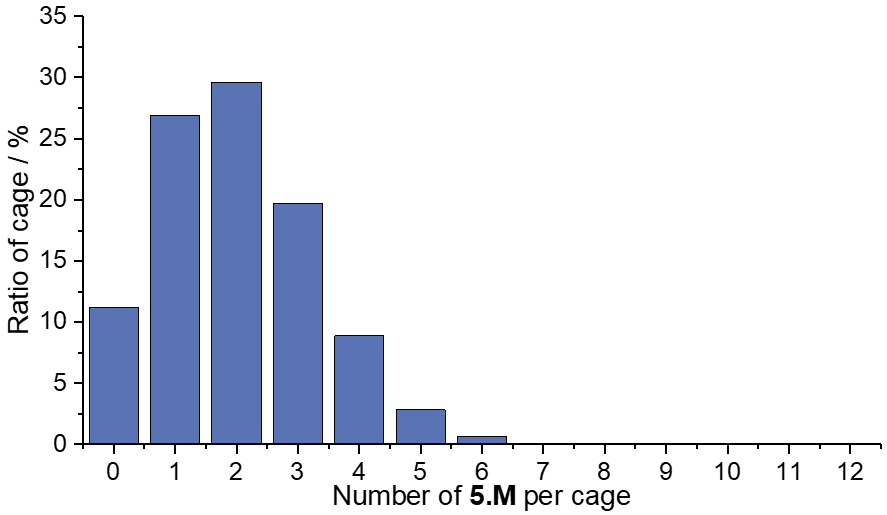


Figure 5.12 | Expected distribution of cages ReO_4^-⊂**5.3** as a function of the number of **5.M** incorporated in each cage.

The UV-vis spectrum of ReO_4^-⊂**5.3** displayed signals corresponding to an MLCT band (maximum absorbance at 552 nm) as well as a band corresponding to dye **5.M** (maximum absorbance at 651 nm). Both bands were slightly red-shifted compared to those of ReO_4^-⊂**5.1** (546 nm) and **5.M** (643 nm), which was attributed to the interaction between the dye and the cage (Figure 5.13 a). The fluorescence spectrum of ReO_4^-⊂**5.3** confirmed the presence of a fluorophore, with the maximum emission wavelength at 661 nm and maximum excitation at 644 nm in water (Figure 5.13 b).

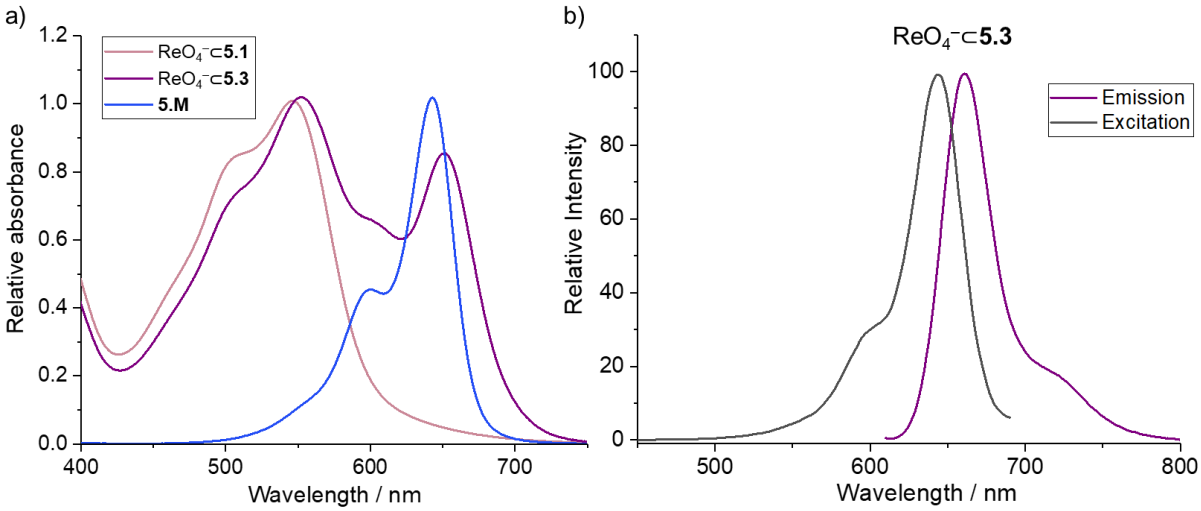


Figure 5.13 | a) UV-vis spectrum of ReO_4^-⊂**5.1**, ReO_4^-⊂**5.3** and **5.M**, showing the presence of similar bands in ReO_4^-⊂**5.3** as in ReO_4^-⊂**5.1** and **5.M**. b) Fluorescence spectra of **5.3** (excitation in black and emission in purple).

The stability of ReO_4^-⊂**5.3** was tested in PBS ([Na_2HPO4] = 10 mM and [NaCl] = 100 mM) using the same method as for ReO_4^-⊂**5.1** in water. To do so, a stock solution of ReO_4^-⊂**5.3** (2.5 mM) was diluted with PBS buffer to different concentrations. The results observed at 0.1 µM, 1 µM and 10 µM were consistent with what was observed for ReO_4^-⊂**5.1** in water, with the only difference that the decomposition followed a linear behaviour from t = 0. We inferred this difference to result from the more controlled pH in this case compared to water. When a higher cage concentration (100 uM) was employed, precipitation of a blue compound, assumed to be **5.M**, was observed, along with an unexpectedly fast decomposition of the cage. This was explained by the low solubility of **5.M** in buffer due to the association of this charged molecule with the different ions present in PBS solution. As one of the subcomponents was coming out of solution, the equilibrium was perturbed, leading to faster decomposition of the cage. When the same experiment was carried in water, no precipitation was observed and the stability of the cage improved. After 14 h, 60% of cage remained for the sample starting at 0.1 µM, 65% for 1 µM and 75% for 10 µM, indicating reasonable stability of ReO_4^-⊂**5.3** in buffer the range 1 µM – 10 µM.

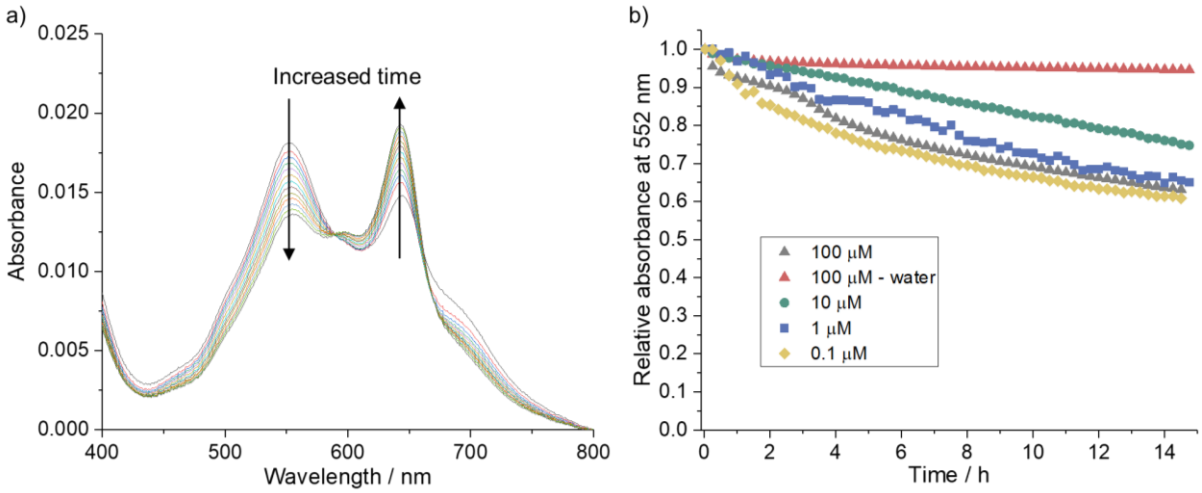


Figure 5.14 | UV-vis spectra in the 400 – 800 nm region corresponding to the MLCT and dye bands of ReO_4^-⊂**5.3** at 10 uM in PBS, showing the changes of the MLCT band (decrease) and dye band (increase) over time (one spectrum every hour). b) Plot of the normalised intensity at 552 nm over time for starting concentrations of ReO_4^- ⊂ **5.3** of 0.1 µM (yellow diamond), 1 µM (blue squares), 10 µM (green circles) and 100 µM (red triangles in water and grey triangle in PBS).

5.4 Uptake of cages ReO_4^-⊂5.3 in hepatocytes

The interaction of ReO_4^-⊂**5.3** with hepatocytes is currently under investigation in the laboratories of Novo Nordisk, Denmark.

5.5 Conclusions and future work

The importance of ASPGR in biological process led us to target this receptor as a way to selectively internalise coordination cages into hepatocytes. Literature precedent showed that the receptor binds strongly to GalNAc residues when arranged in a triangular fashion. In this chapter, a modified picolinaldehyde subcomponent was therefore synthesised with the aim to bioconjugate the GalNac targeting moiety onto self-assembled cages. The design of this subcomponent was carefully chosen such that, once self-assembled, each metal vertex of the cage would present three targeting units, arranged optimally in space to bind to the ASGPR.

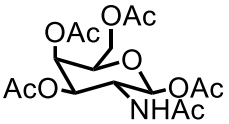
Self-assembly of the bioconjugated subcomponent was successful and water soluble $Fe^{II}_4L_4$ cages were obtained. The coordination cages were further functionalised with a fluorescent tag to facilitate their tracking during cell studies. To do so, some of the GalNAc units were replaced by an analogue of the Cy5 dye on the coordination cage. The attachment of both the GalNAc moieties and the fluorescent tag onto the cage was shown by a combination of NMR, DOSY and UV-vis experiments. The stability of the cage formed was tested in biological conditions (PBS) and was shown to only slowly decompose to a small extent when the concentration was kept between 1 µM and 10 µM. This range of concentrations were found suitable for biological testing and the interaction of the cage formed with hepatocytes is being investigated.

Future work will build upon this proof of concept by extending this strategy of bioconjugation to other cages, which display better stability and lower toxicity. The use of different coordination motifs such as bipyridine[13] or pyridyl-triazole[14] could be tested instead of the pyridyl-imine used in this chapter for example. An obvious next step is the use of the cage cavity to carry cargoes into cells. By careful design of the ligands, drugs of interest could be encapsulated and delivered selectively to specific cells. Incorporation of responsive groups onto the cage could also lead to controlled release of the cargo upon application of stimuli.

5.6 Experimental section

5.6.1 Synthesis of 5.J

2-Acetamido-2-deoxy-1,3,4,6-tetra-O-acetyl-β-D-galactopyranoside 5.A



D-galactosamine (10.0 g, 55.8 mmol, 1.00 eq.) was suspended in anhydrous pyridine (100 mL). Acetic anhydride (60.0 mL, 635 mmol, 11.4 eq.) was added dropwise over 1 h. The mixture was stirred at room temperature for 18 h, after which the solvents were removed under reduced pressure, giving a syrupy transparent liquid. Addition of CH_3OH (100 mL) precipitated the product which was collected by filtration and washed with CH_3OH (200 mL), giving **5.A** as a white powder (20.6 g, 53.0 mmol, 95%).

^{1}H NMR (400 MHz, 298 K, $CDCl_3$) δ 5.70 (d, *J* = 9.2 Hz, 1H), 5.41 (d, *J* = 9.2 Hz, 1H), 5.37 (dd, *J* = 3.4, 1.2 Hz, 1H), 5.08 (dd, *J* = 11.3, 3.4 Hz, 1H), 4.45 (dt, *J* = 11.3, 9.2 Hz, 1H), 4.24 – 4.05 (m, 2H), 4.02 (td, *J* = 6.5, 1.2 Hz, 1H), 2.17 (s, 3H), 2.13 (s, 3H), 2.05 (s, 3H), 2.02 (s, 3H), 1.94 (s, 3H).

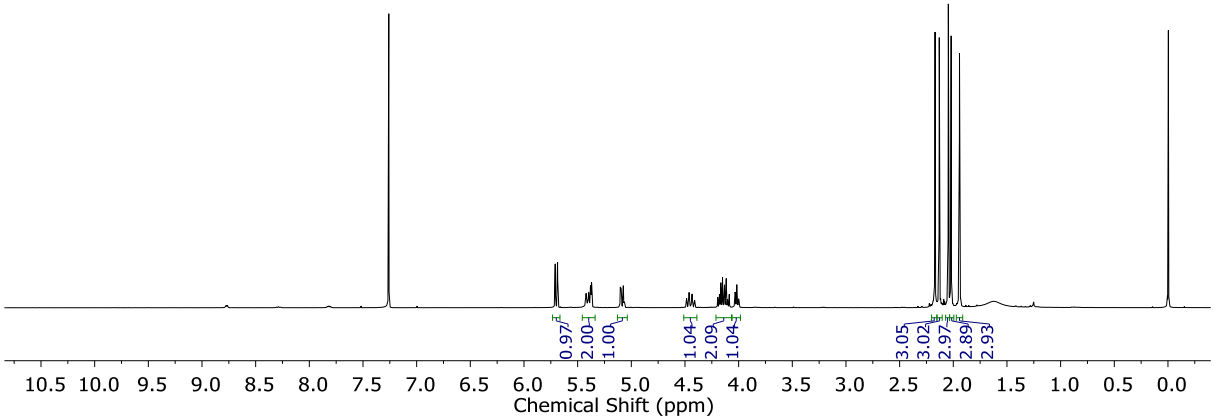
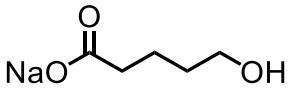


Figure 5.15 | ^{1}H NMR spectrum (400 MHz, 298 K, $CDCl_3$) of **5.A.**

This data was consistent with previously reported data.[15]

Sodium 5-hydroxypentanoate 5.B



δ-Valerolactone (15.0 g, 150 mmol, 1.00 eq.) and NaOH (6.00 g, 150 mmol, 1.00 eq.) were dissolved in H_2O (150 mL) and stirred at 70 °C for 18 h. The mixture was cooled to room temperature and the solvents removed under reduced pressure to give **5.B** as a white solid (20.6 g, 147 mmol, 98%).

^{1}H NMR (400 MHz, 298 K, D_2O) δ 3.62 (t, *J* = 6.3 Hz, 2H), 2.21 (t, *J* = 7.2 Hz, 2H), 1.75 – 1.39 (m, 4H).

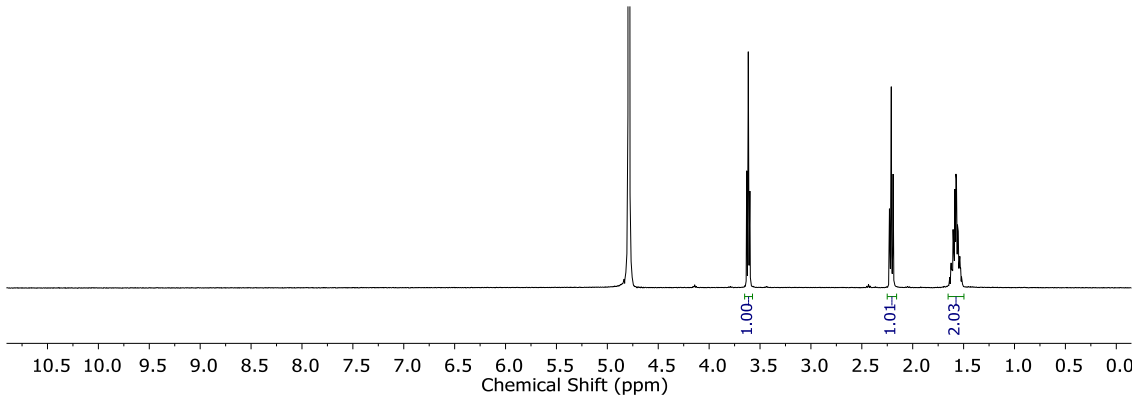
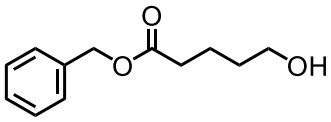


Figure 5.16 | ^{1}H NMR spectrum (400 MHz, 298 K, D_2O) of **5.B.**

This data was consistent with previously reported data.[16]

Benzyl 5-hydroxypentanoate 5.C



Sodium 5-hydroxypentanoate (9.50 g, 67.8 mmol, 1.00 eq.) was suspended in acetone (70 mL) and sonicated for 3 h. Benzyl bromide (9.6 mL, 80.7 mmol, 1.19 eq.) and TBABr (1.10 g, 3.41 mmol, 0.05 eq.) were added and the mixture was refluxed for 18 h. The solvents was removed under reduced pressure and the resulting oil was dissolved in EtOAc (100 mL) and washed with an aqueous saturated solution of $NaHCO_3$ (3 × 100 mL) and brine (3 × 100 mL). The organic phase was dried with $MgSO_4$ and the solvent was removed under reduced pressure to give a light yellow oil. The product was purified by silica gel chromatography (EtOAc in heptane, 0 – 50%) to give **5.C** as a yellow oil (4.37 g, 21.0 mmol, 31%).

^{1}H NMR (400 MHz, 298 K, $CDCl_3$) δ 7.43 – 7.29 (m, 5H), 5.12 (s, 2H), 3.64 (t, *J* = 6.4 Hz, 2H), 2.41 (t, *J* = 7.3 Hz, 2H), 1.83 – 1.68 (m, 2H), 1.66 – 1.52 (m, 3H).

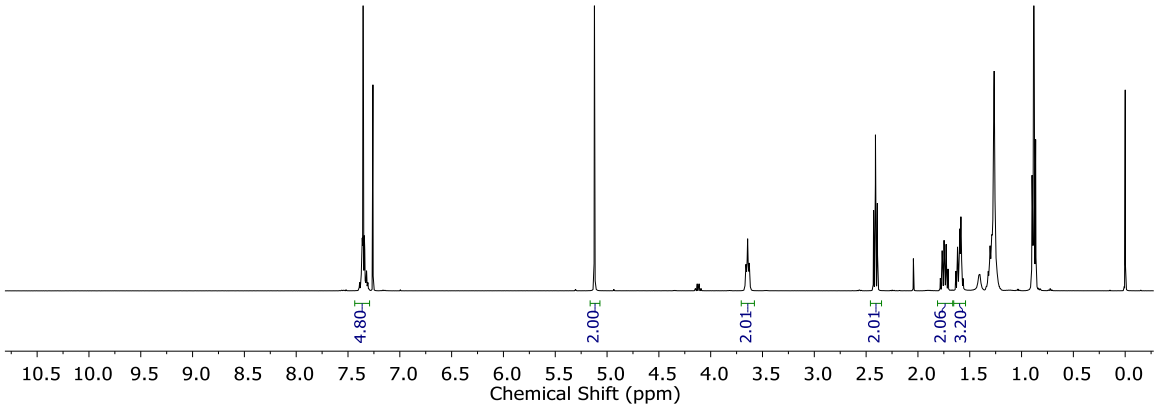
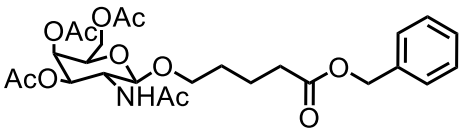


Figure 5.17 | ^{1}H NMR spectrum (400 MHz, 298 K, $CDCl_3$) of **5.C**. Peaks for residual heptane were observed.

This data was consistent with previously reported data.[17]

(2R,3R,4R,5R,6R)-5-Acetamido-2-(acetoxymethyl)-6-((5-(benzyloxy)-5-oxopentyl)oxy)tetrahydro-2H-pyran-3,4-diyl diacetate 5.D



Compound **5.A** (1.15 g, 2.95 mmol, 1.00 eq.), benzyl 5-hydroxypentanoate **5.C** (1.22 g, 5.86 mmol, 2.00 eq.) and $Sc(OTf)_3$ (0.10 g, 0.20 mmol, 0.07 eq.) were combine in DCE (10 mL) and refluxed under nitrogen for 3 h. The mixture was cooled to room temperature, diluted with DCM (100 mL) and washed with an aqueous saturated solution of $NaHCO_3$ (3 × 100 mL). The organic phase was dried with $MgSO_4$ and the solvents were removed under reduced pressure. The product was purified by prep-HPLC (C18 column, 0.1% TFA, CH_3CN in H_2O, 10 – 65%) to give **5.D** as a yellow oil (0.238 g, 0.442 mmol, 15%).

^{1}H NMR (400 MHz, 298 K, $CDCl_3$) δ 7.45 – 7.30 (m, 5H), 5.79 (d, *J* = 8.7 Hz, 1H), 5.36 (dd, *J* = 3.4, 1.1 Hz, 1H), 5.26 (dd, *J* = 11.2, 3.4 Hz, 1H), 5.13 (s, 2H), 4.64 (d, *J* = 8.4 Hz, 1H), 4.24 – 4.08 (m, 2H), 3.98 (dt, *J* = 11.2, 8.5 Hz, 1H), 3.93 – 3.83 (m, 2H), 3.51 (ddd, *J* = 9.9, 6.7, 5.5 Hz, 1H), 2.48 – 2.31 (m, 2H), 2.16 (s, 3H), 2.06 (s, 3H), 2.01 (s, 3H), 1.94 (s, 3H), 1.81 – 1.67 (m, 2H), 1.68 – 1.54 (m, 2H).

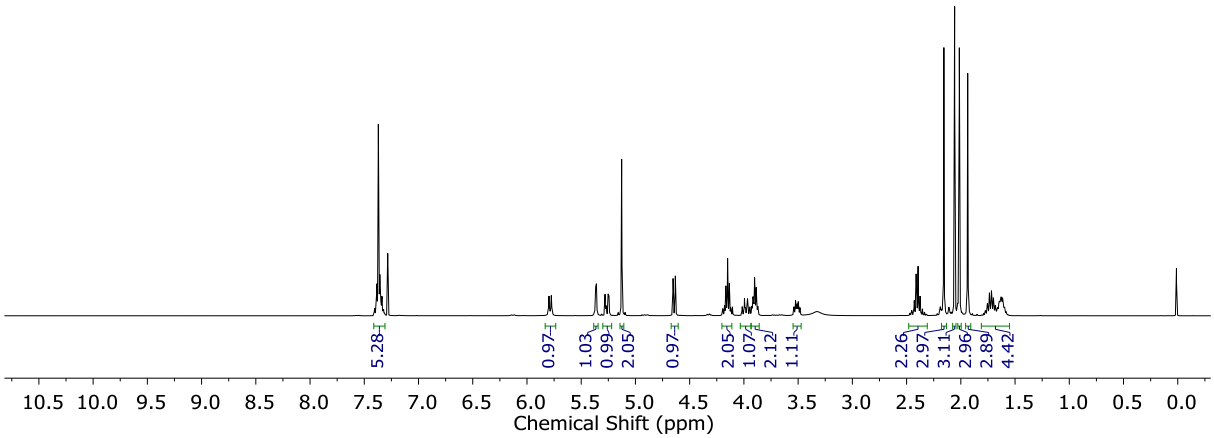
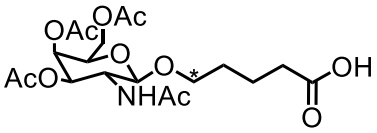


Figure 5.18 | ^{1}H NMR spectrum (400 MHz, 298 K, $CDCl_3$) of **5.D.**

This data was consistent with previously reported data.[18]

5-(((2R,3R,4R,5R,6R)-3-Acetamido-4,5-diacetoxy-6-(acetoxymethyl)tetrahydro-2H-pyran-2-yl)oxy)pentanoic acid 5.E



Compound **5.D** (100 mg, 0.186 mmol) was dissolved in THF (20 mL) and Pd/C (10 mg) was added. The mixture was stirred under a hydrogen atmosphere for 18 h. The Pd/C was removed

by filtration on Celite and the solvent removed under reduced pressure to give the product as an off-white solid (79.8 mg, 0.179 mmol, 96%).

^{1}H NMR (400 MHz, 298 K, DMSO-d_6) δ 7.82 (d, J = 9.2 Hz, 1H), 5.22 (d, J = 3.4 Hz, 1H), 4.96 (dd, J = 11.2, 3.5 Hz, 1H), 4.49 (d, J = 8.5 Hz, 1H), 4.06 – 3.99 (m, 2H), 3.88 (dt, J = 11.3, 8.9 Hz, 1H), 3.77 – 3.65 (m, 1H), 2.25 – 2.13 (m, 2H), 2.11 (s, 3H), 2.00 (s, 3H), 1.90 (s, 3H), 1.77 (s, 3H), 1.58 – 1.42 (m, 4H). The protons marked by the asterisk were inferred to be under the water peak.

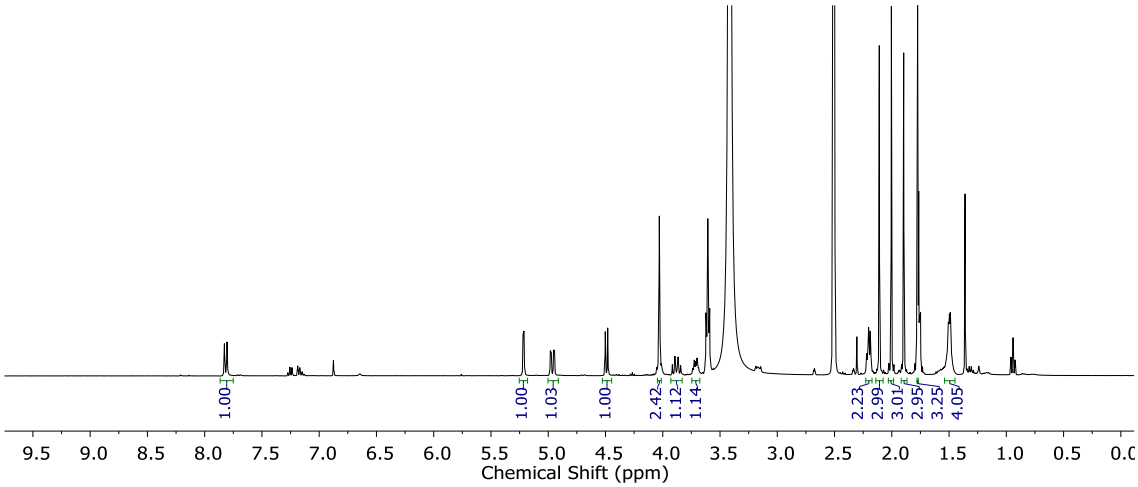
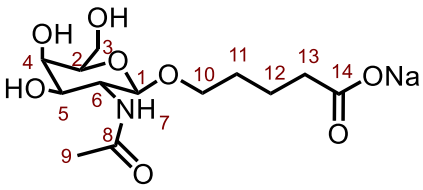


Figure 5.19 | ^{1}H NMR spectrum (400 MHz, 298 K, DMSO-d_6) of **5.E**. Peaks for residual THF were observed.

This data was consistent with previously reported data.[18]

5-(((2R,3R,4R,5R,6R)-3-Acetamido-4,5-dihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)pentanoic acid 5.F



Compound **5.E** (79.8 mg, 0.179 mmol) was dissolved in a solution of sodium methoxide (1 M) in CH_3OH (2 mL) and stirred at room temperature for 18 h. The solvent was removed under reduced pressure. The product **5.F** was obtained as a white solid (52.9 mg, 0.165 mmol, 92%).

^{1}H NMR (400 MHz, 298 K, CD_3OD) δ 4.34 (d, J = 8.4 Hz, 1H, H_1), 3.96 – 3.83 (m, 3H, $H_{6,10}$), 3.77 (d, J = 5.8 Hz, 2H, H_3), 3.61 – 3.49 (m, 2H, $H_{5,4}$), 3.46 (t, J = 5.8 Hz, 1H, H_2), 2.15 (t, J = 7.1 Hz, 2H, H_{13}), 1.98 (s, 3H, H_9), 1.69 – 1.53 (m, 4H, $H_{12,11}$). **^{13}C NMR** (101 MHz, 298 K, CD_3OD) δ 181.0, 172.5, 101.3, 74.7, 71.7, 68.6, 68.4, 60.9, 52.4, 36.9, 28.7, 22.2, 21.3. **HRMS**: *m/z* calculated for [**5.F** + H^+] = 344.1321, observed = 344.1400.

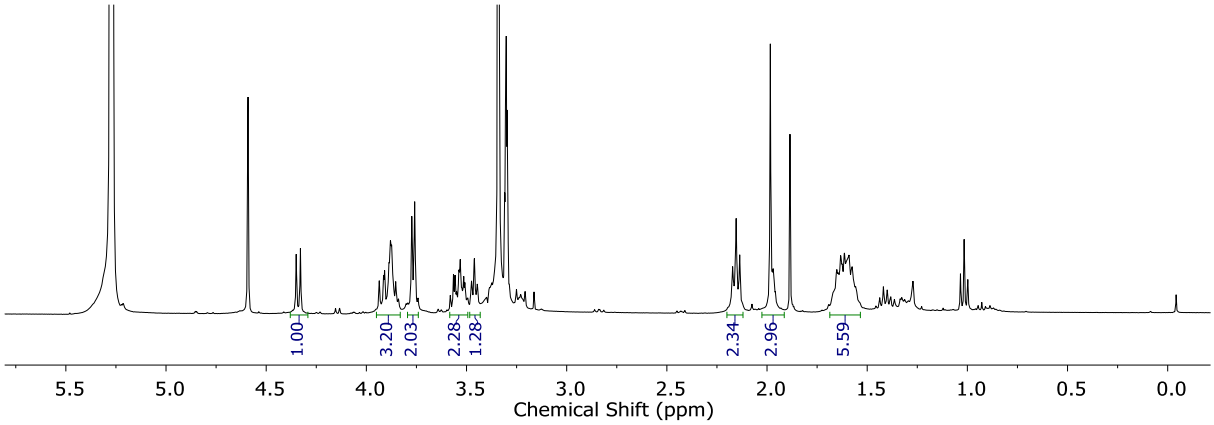


Figure 5.20 | ^{1}H NMR spectrum (400 MHz, 298 K, CD_3OD) of **5.F.** Peaks for residual toluene and TBA are also observed (4.60, 3.23, 1.41 and 1.01 ppm).

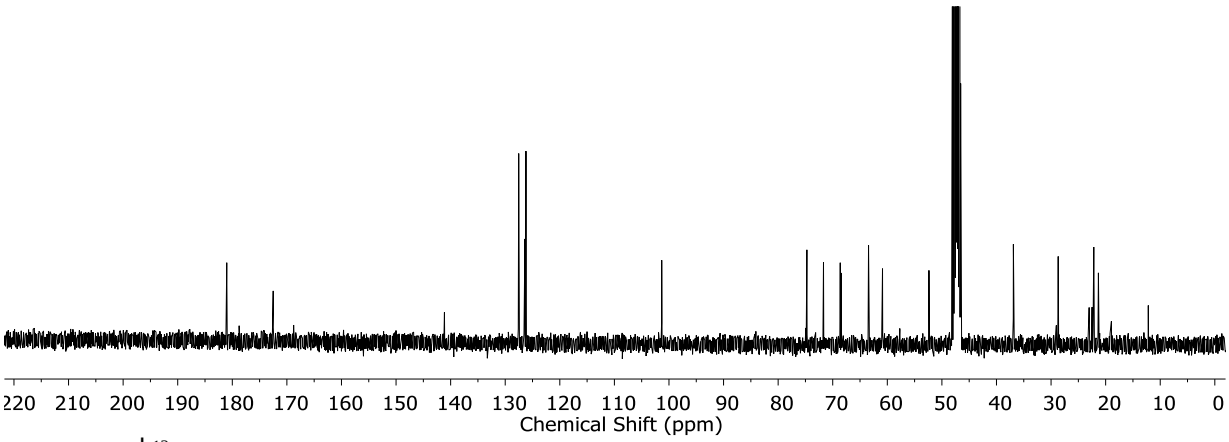


Figure 5.21 | ^{13}C NMR spectrum (101 MHz, 298 K, CD_3OD) of **5.F**. Peaks for residual toluene are also observed (127-126 and 63 ppm).

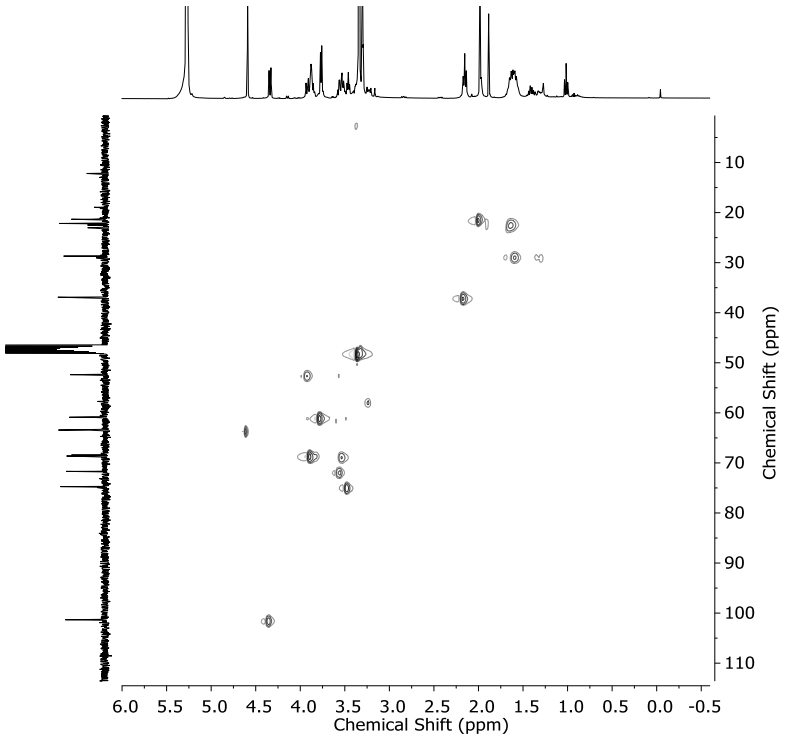


Figure 5.22 | ^{1}H-^{13}C HSQC spectrum (400 MHz, 298 K, CD_3OD) of **5.F.**

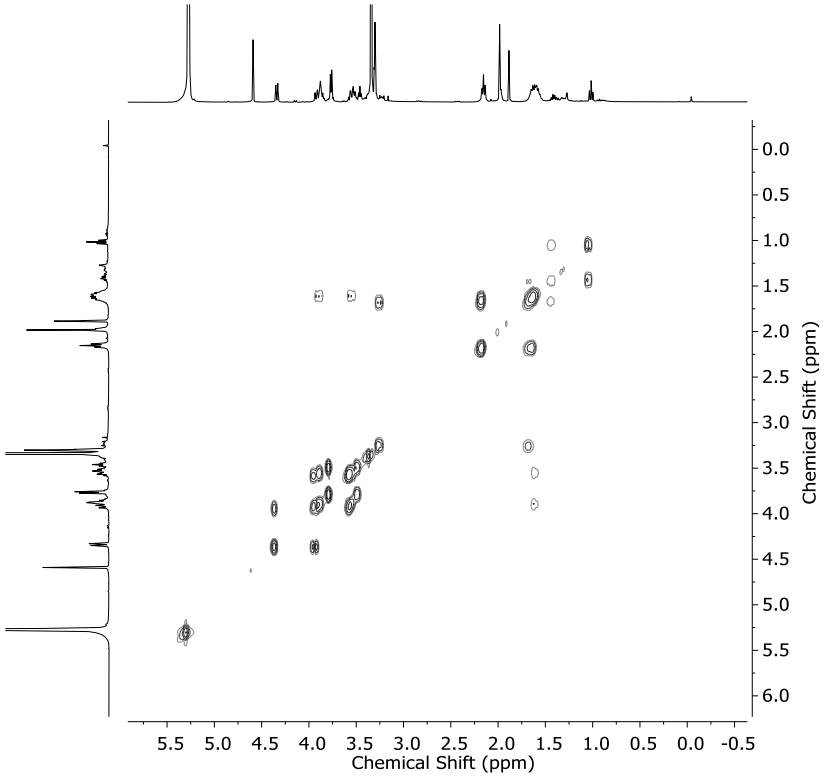
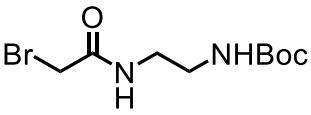


Figure 5.23| ^{1}H-^{1}H COSY spectrum (400 MHz, 298 K, CD_3OD) of **5.F.**

tert-Butyl (2-(2-bromoacetamido)ethyl)carbamate 5.G



2-Bromoacetyl bromide (132 µL, 1.52 mmol, 1.20 eq.) was dissolved in DCM (20 mL) and combined with K_2CO_3 (202 mg, 1.60 mmol, 1.30 eq.) dissolved in water (20 mL). *tert*-Butyl (2-aminoethyl)carbamate (200 µL, 1.26 mmol, 1.00 eq.) was dissolved in DCM (50 mL) and added to the solution. The mixture was stirred at room temperature for 1.5 h after which the organic layer was separated, washed with H_2O (3 × 50 mL) and brine (3 × 50 mL). The organic layer was dried with $MgSO_4$ and the solvent was removed under reduced pressure to give **5.G** as a white powder (260 mg, 0.925 mmol, 71%).

^{1}H NMR (400 MHz, 298 K, $CDCl_3$) δ 7.10 (bs, 1H), 4.88 (bs, 1H), 3.86 (s, 2H), 3.39 (q, *J* = 5.5 Hz, 2H), 3.31 (q, *J* = 5.5 Hz, 2H), 1.45 (s, 9H).

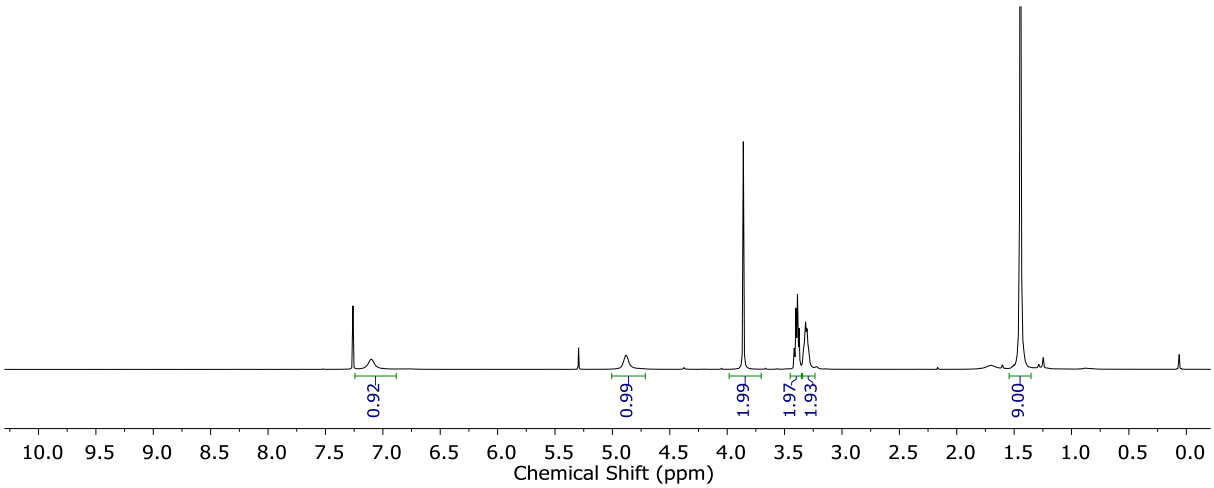
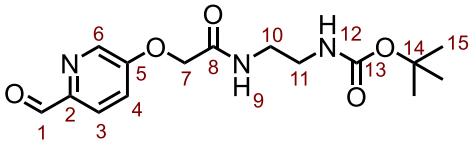


Figure 5.24 | ^{1}H NMR spectrum (400 MHz, 298 K, $CDCl_3$) of **5.G.**

This data was consistent with previously reported data.[19]

***tert-Butyl (2-(2-((6-formylpyridin-3-yl)oxy)acetamido)ethyl)carbamate 5*.H**



Compound **5.G** (240 mg, 0.854 mmol, 1.00 eq.), 5-hydroxypicolinaldehyde (116 mg, 0.943 mmol, 1.10 eq.) and DIPEA (163 µL, 0.943 mmol, 1.10 eq.) were combined in DMF (2.5 mL) and stirred at 100 °C for 2 h. The mixture was diluted with H_2O (10 mL), extracted with EtOAc (3 × 20 mL), dried with $MgSO_4$ and the solvents were removed under reduced pressure. The product was purified by silica gel chromatography (EtOAc in hexane, 60-100%) to give **5.H** as a white crystalline solid (179 mg, 0.555 mmol, 65%).

^{1}H NMR (400 MHz, 298 K, $CDCl_3$) δ 9.99 (s, 1H, H_1), 8.52 (d, *J* = 2.8 Hz, 1H, H_6), 7.98 (d, *J* = 8.6 Hz, 1H, H_3), 7.56 (bs, 1H, H_9), 7.35 (dd, *J* = 8.6, 2.8 Hz, 1H, H_4), 4.95 (bs, 1H, H_{12}), 4.60 (s, 2H, H_7), 3.55 – 3.39 (m, 2H, H_{10}), 3.41 – 3.26 (m, 2H, H_{11}), 1.41 (s, 9H, H_{15}). **^{13}C NMR** (101 MHz, 298 K, $CDCl_3$) δ 191.4, 166.7, 156.9, 156.2, 146.9, 138.6, 123.0, 120.6, 79.7, 67.0, 40.9, 39.5, 27.9. **HRMS**: *m/z* calculated for [**5.H** + H^+] = 324.1554, observed = 324.1557

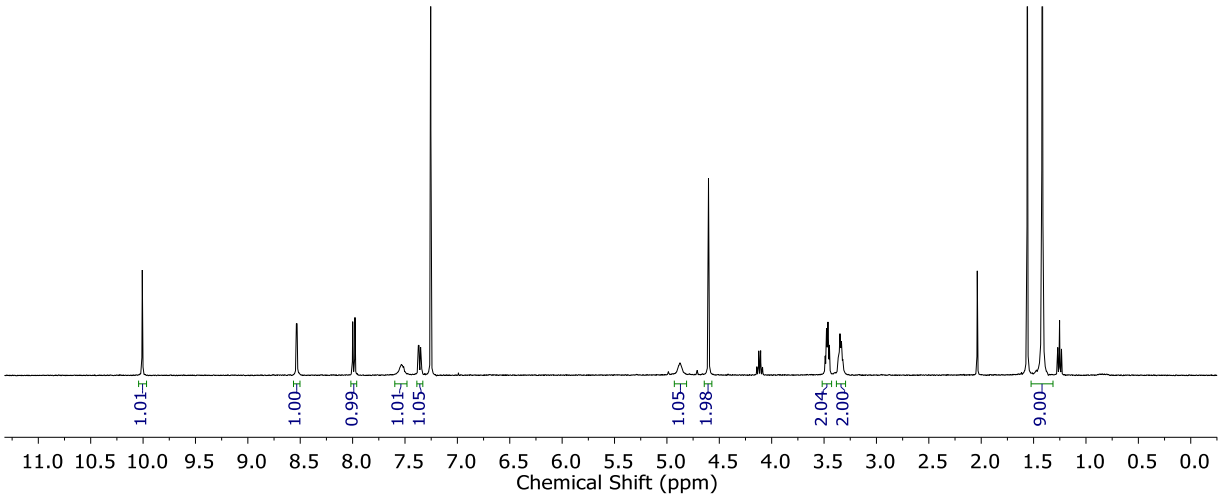


Figure 5.25 | ^{1}H NMR spectrum (400 MHz, 298 K, $CDCl_3$) of **5.H.**

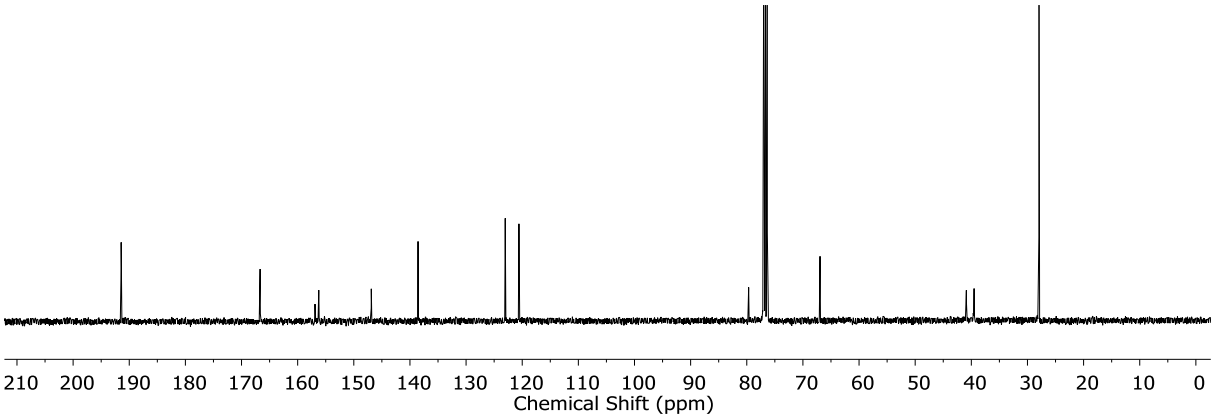


Figure 5.26 | ^{13}C NMR (101 MHz, 298 K, $CDCl_3$) of **5.H**.

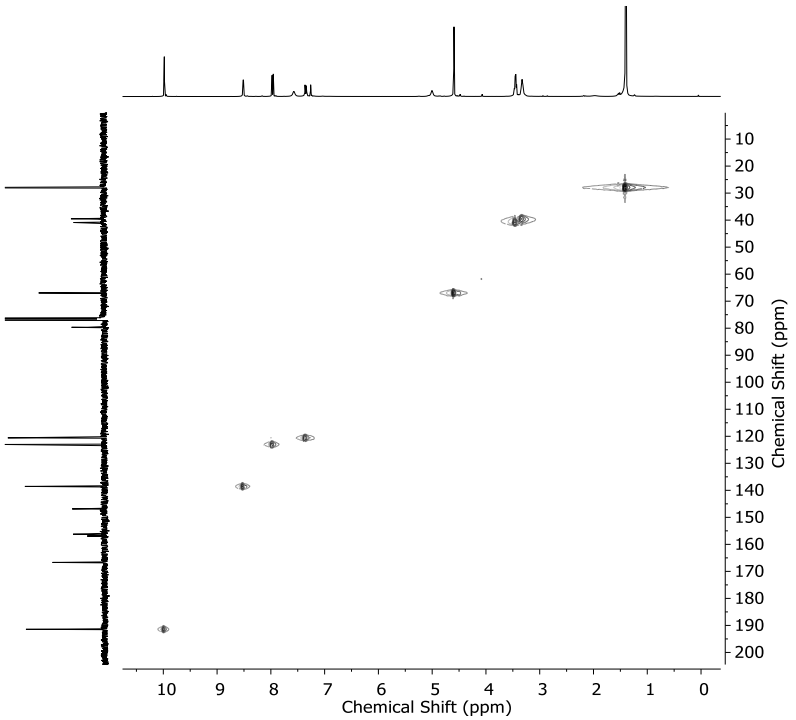


Figure 5.27 | ^{1}H-^{13}C HSQC spectrum (400 MHz, 298 K, $CDCl_3$) of **5.H.**

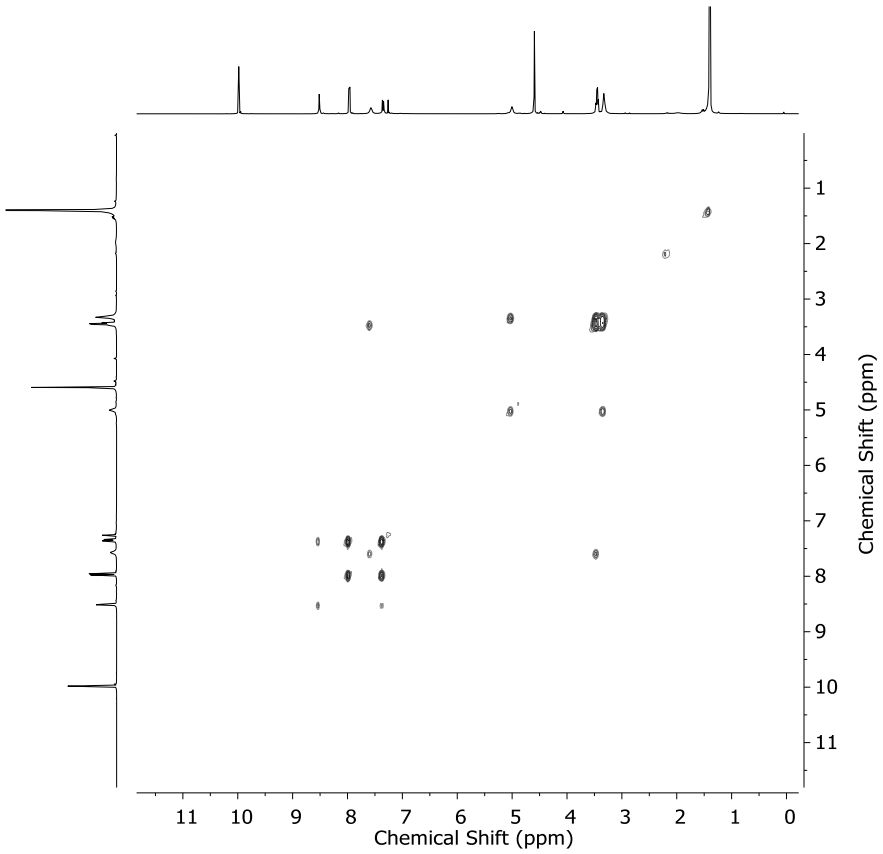
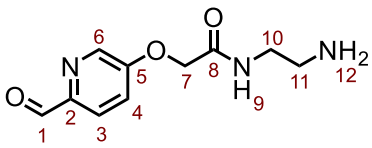


Figure 5.28 | ^{1}H-^{1}H COSY spectrum (400 MHz, 298 K, $CDCl_3$) of **5.H.**

N-(2-Aminoethyl)-2-((6-formylpyridin-3-yl)oxy)acetamide 5.I



Compound **5.H** (100 mg, 0.309 mmol) was dissolved in 1,4-dioxane (1.5 mL) and HCl/dioxane (4 M, 2.5 mL) were added. The solution was stirred for 5 min after which the solid was separated by centrifugation. The solid was washed with 1,4-dioxane, DCM, EtOAc, hexane and finally Et_2O. The hydroscopic solid was freeze-dried to give **5.I** as an off-white solid (68.2 mg, 0.306 mmol, 99%).

The compound was hypothesized to be the hydrochloride adduct due to protonation of both pyridine and amine moieties.

^{1}H NMR (400 MHz, 298 K, DMSO-d_6) δ 9.90 (s, 1H, H_1), 8.56 (d, J = 2.8 Hz, 1H, H_6), 8.51 (bt, J = 5.8 Hz, 1H, H_9), 8.00 (bs, 2H, H_{12}), 7.95 (d, J = 8.7 Hz, 1H, H_3), 7.62 (dd, J = 8.7, 2.8 Hz, 1H, H_4), 4.77 (s, 2H, H_7), 3.40 (q, J = 6.1 Hz, 2H, H_{10}), 2.90 (q, J = 6.1 Hz, 2H, H_{11}). **^{13}C NMR** (101 MHz, 298 K, DMSO-d_6) δ 192.0, 167.5, 157.5, 146.0, 139.0, 123.7, 121.8, 67.1, 38.5, 36.3. **HRMS**: *m/z* calculated for [**5.I** + H^+] = 224.1030, observed = 224.1028.

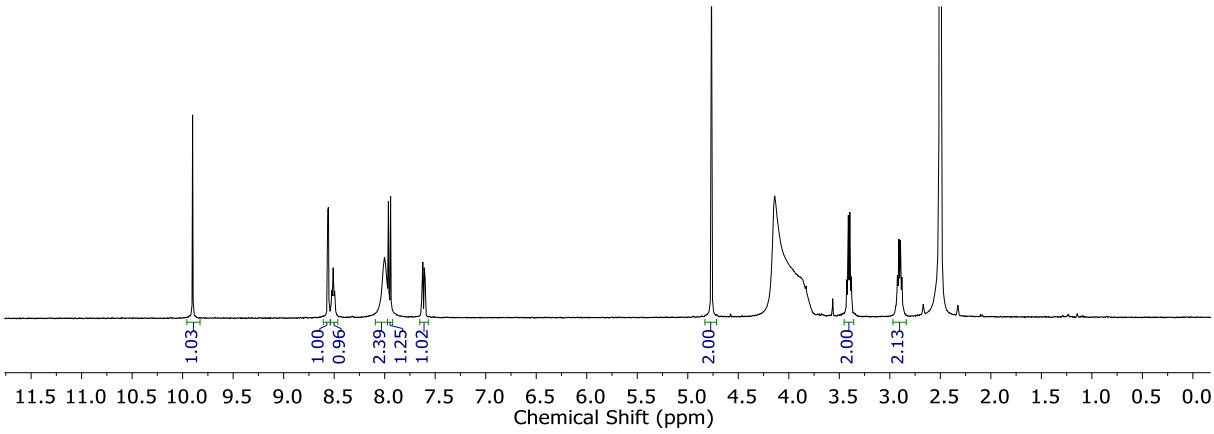


Figure 5.29 | ^{1}H NMR spectrum (400 MHz, 298 K, DMSO-d_6) of **5.I.**

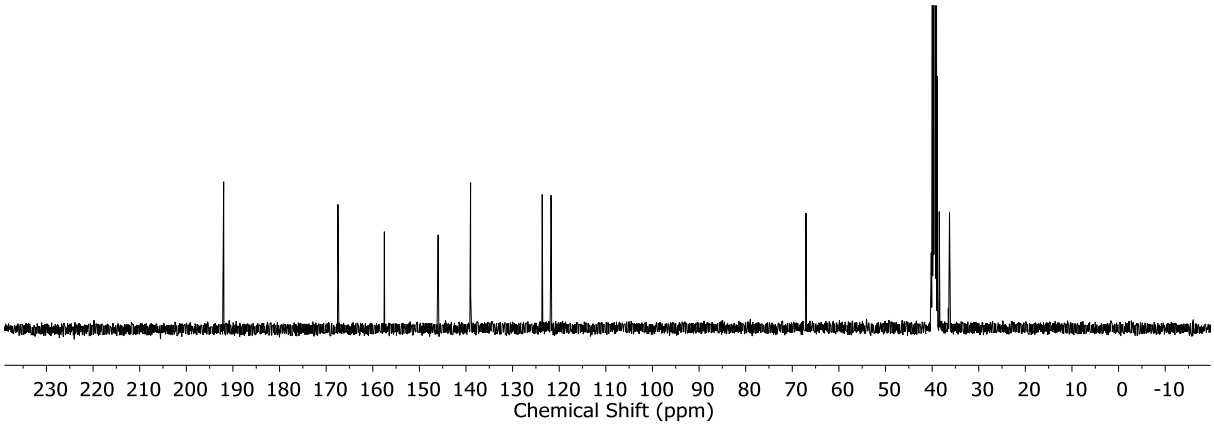


Figure 5.30 | ^{13}C NMR (101 MHz, 298 K, DMSO-d_6) of **5.I**.

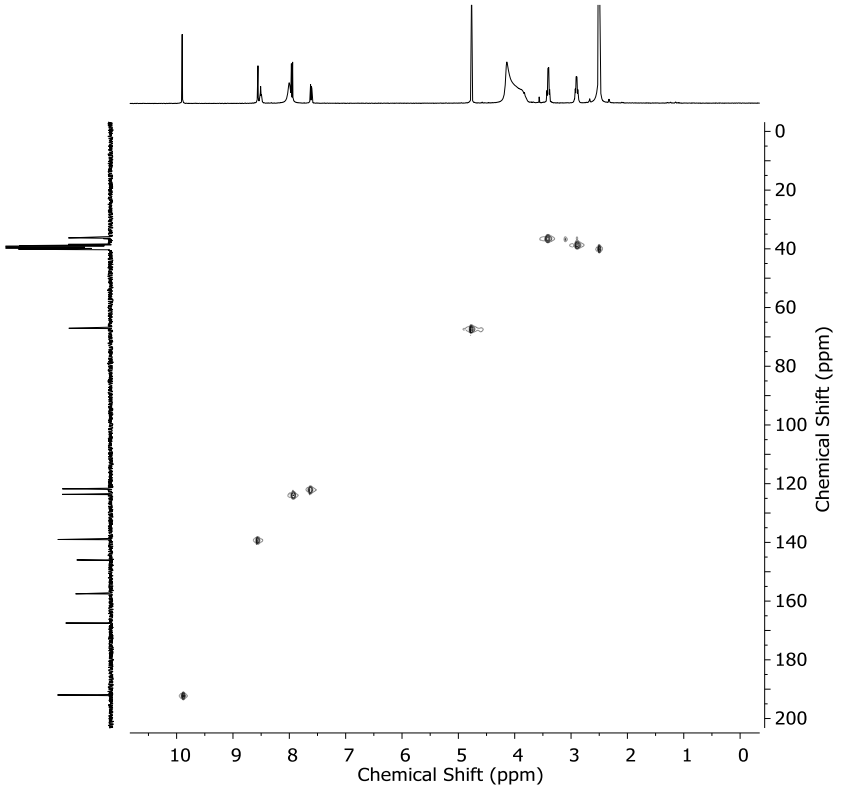


Figure 5.31 | ^{1}H-^{13}C HSQC spectrum (400 MHz, 298 K, DMSO-d_6) of **5.I.**

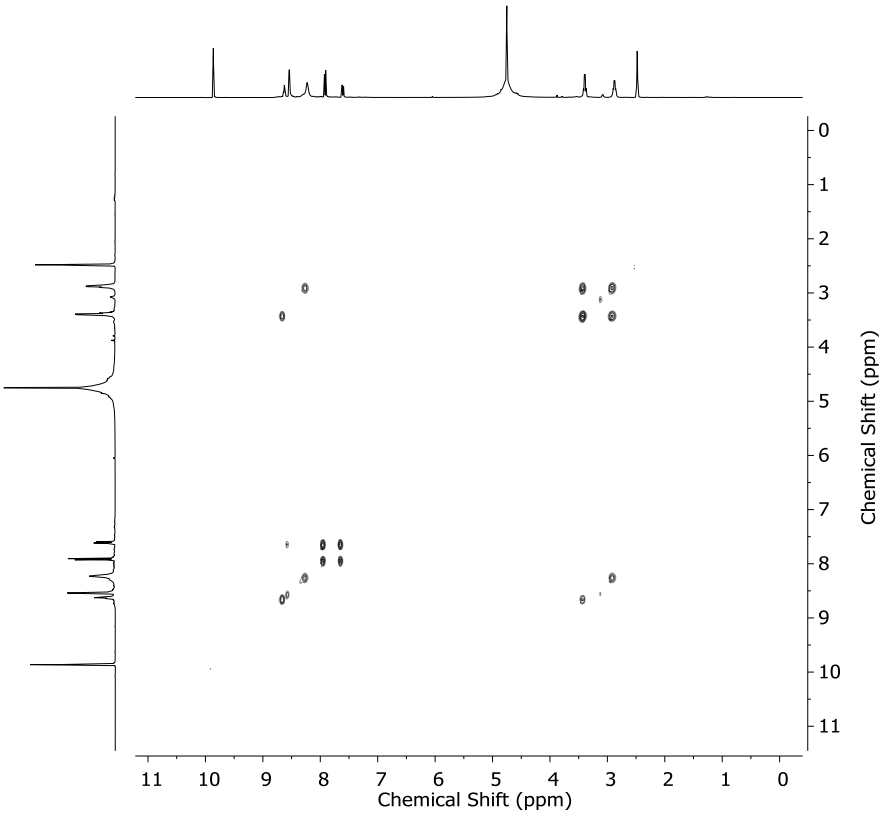
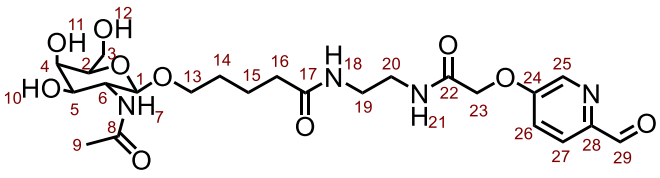


Figure 5.32| ^{1}H-^{1}H COSY spectrum (400 MHz, 298 K, DMSO-d_6) of **5.I.** Change in the amide proton shifts was attributed to a change in the acidity and water content of the solvent (as evidenced by the shift of the water peak).

5-(((2R,3R,4R,5R,6R)-3-Acetamido-4,5-dihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)-N-(2-(2-((6-formylpyridin-3-yl)oxy)acetamido)ethyl)pentanamide 5.J



Compound **5.I** (36.9 mg, 0.124 mmol, 1.00 eq.), compound **5.F** (40.0 mg, 0.124 mmol, 1.00 eq.), HBTU (47.2 mg, 0.124 mmol, 1.00 eq.) and DIPEA (172 µL, 0.992 mmol, 8.00 eq.) were combined in DMF (2 mL) and stirred at room temperature for 2 h. The product was precipitated with Et_2O (10 mL), collected by centrifugation and washed with DCM, EtOAc and finally Et_2O, giving **5.J** as a white powder (60.7 mg, 0.115 mmol, 93%).

^{1}H NMR (400 MHz, 298 K, DMSO-d_6) δ 9.90 (s, 1H, H_{29}), 8.52 (d, J = 2.8 Hz, 1H, H_{25}), 8.27 (t, J = 6.2 Hz, 1H, H_{21}), 7.95 (d, J = 8.6 Hz, 1H, H_{27}), 7.84 (t, J = 5.6 Hz, 1H, H_{18}), 7.61 (d, J = 9.0 Hz, 1H, H_7), 7.56 (dd, J = 8.7, 2.9 Hz, 1H, H_{26}), 4.71 (s, 2H, H_{23}), 4.61 – 4.51 (m, 2H, $H_{11,12}$), 4.46 (d, J = 4.3 Hz, 1H, H_{10}), 4.21 (d, J = 8.4 Hz, 1H, H_1), 3.75 – 3.60 (m, 2H, $H_{6,4}$), 3.60 – 3.45 (m, 2H, H_3), 3.43 – 3.39 (m, 1H, H_5), 3.29 – 3.24 (m, 1H, H_2), 3.22 – 3.10 (m, 4H,

$H_{19,20}$), 2.04 (t, J = 7.2 Hz, 2H, H_{16}), 1.79 (s, 3H, H_9), 1.56 – 1.36 (m, 4H, $H_{14,15}$). H_{13} was found under the water peak (by COSY and HSQC). **^{13}C NMR** (101 MHz, 298 K, DMSO-d_6) δ 192.1, 172.6, 169.7, 166.9, 157.5, 146.1, 139.1, 123.6, 121.5, 101.5, 75.3, 71.6, 68.0, 67.6, 60.5, 55.3, 52.2, 38.6, 38.1, 35.2, 28.7, 23.1, 21.9. **HRMS**: *m/z* calculated for [**5.J** + Na^+] = 549.2167, observed = 549.2169.

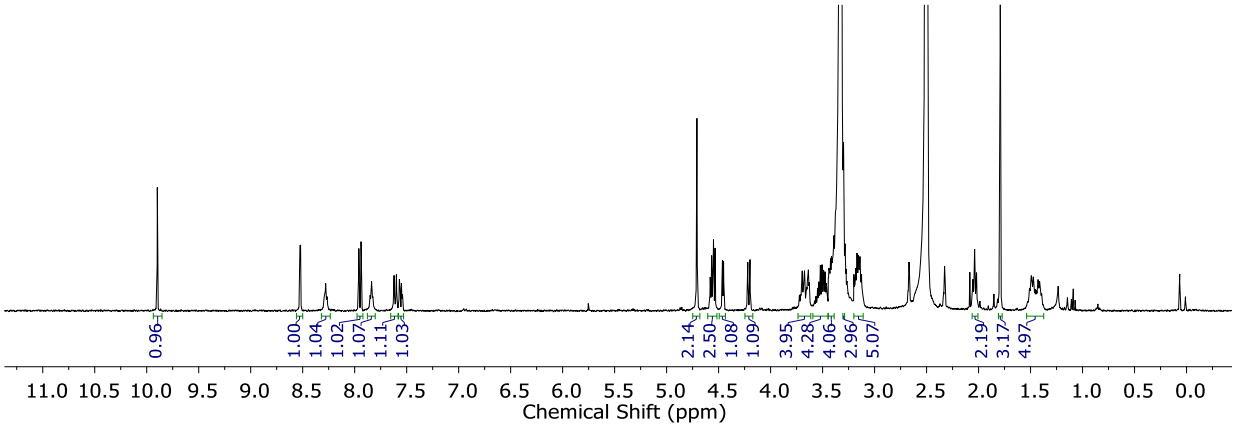


Figure 5.33| ^{1}H NMR spectrum (400 MHz, 298 K, DMSO-d_6) of **5.J.**

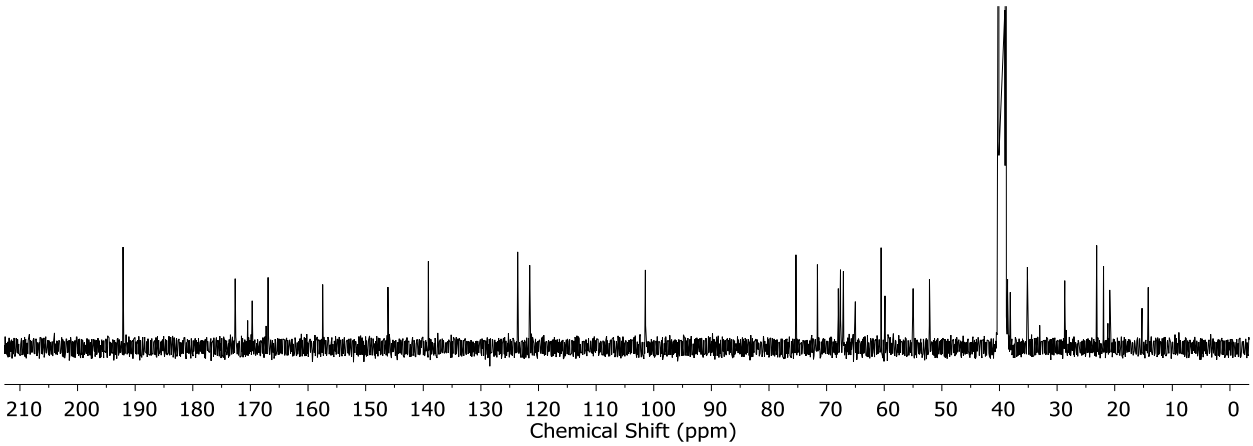


Figure 5.34| ^{13}C NMR (101 MHz, 298 K, DMSO-d_6) of **5.J**.

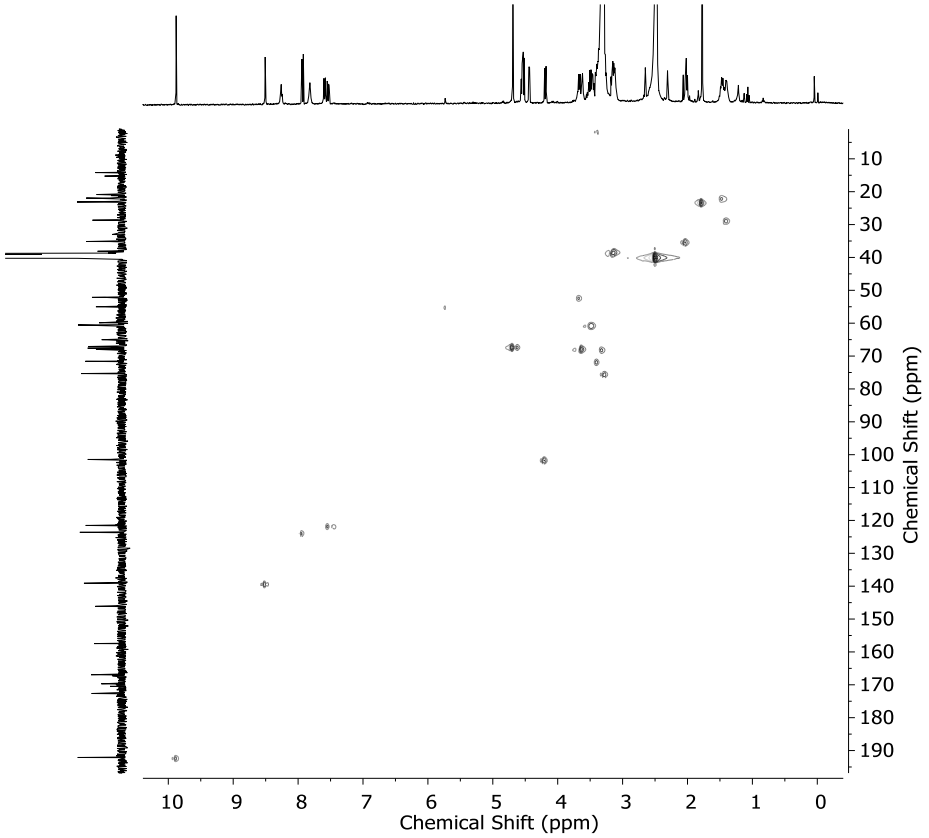


Figure 5.35| ^{1}H-^{13}C HSQC spectrum (400 MHz, 298 K, DMSO-d_6) of **5.J.**

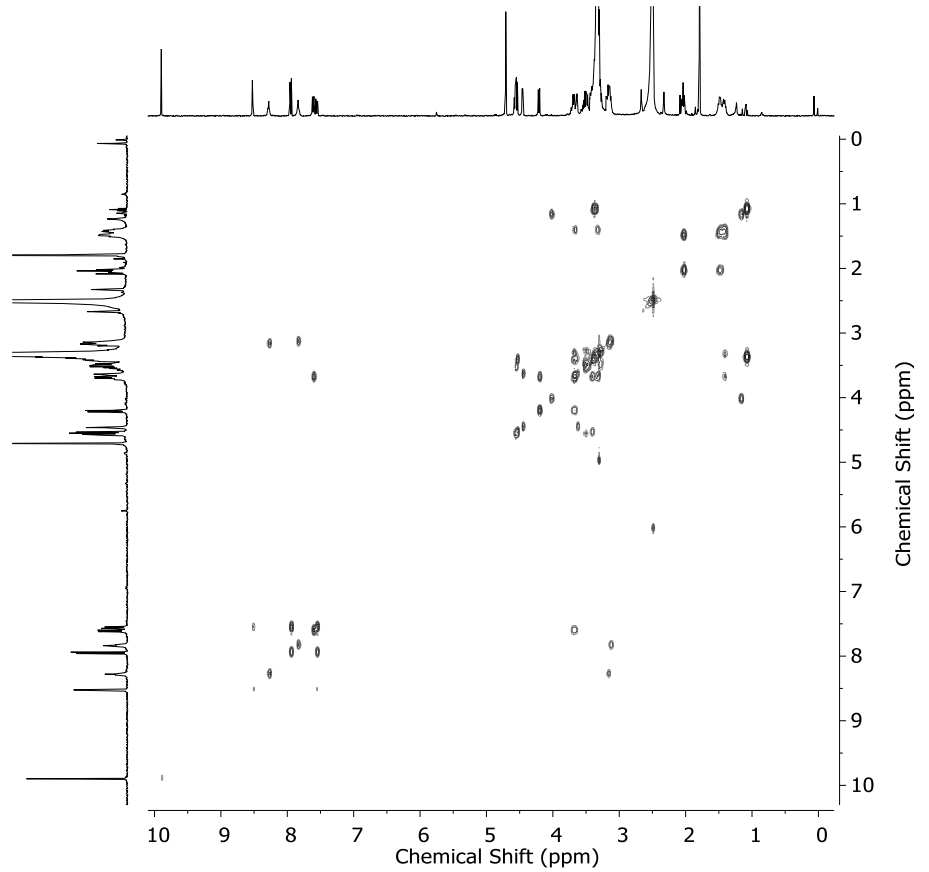
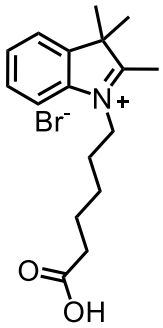


Figure 5.36| ^{1}H-^{1}H COSY spectrum (500 MHz, 298 K, DMSO-d_6) of **5.J.**

5.6.2 Synthesis of 5.M

Compound **5.K** and **5.L** were synthesized from an adapted literature procedure.[10]

1-(5-Carboxypentyl)-2,3,3-trimethyl-3H-indol-1-ium bromide 5.K.



2,3,3-Trimethyl-3*H*-indole (2.00 g, 6.30 mmol, 1.00 eq.) and 6-bromohexanoic acid (2.45 g, 6.30 mmol, 1.00 eq.) were combined in CH_3NO_2 (8.0 mL). The mixture was stirred at 80 °C for 6 h. Addition of Et_2O resulted in the precipitation of a dark red paste which was further triturated with Et_2O to give **5.K** as a dark red solid (1.21 g, 3.42 mmol, 27%).

^{1}H NMR (400 MHz, 298 K, DMSO-d_6) δ 7.96 (dd, J = 6.0, 2.9 Hz, 1H), 7.83 (dd, J = 6.0, 2.9 Hz, 1H), 7.70 – 7.50 (m, 2H), 4.44 (t, J = 7.8 Hz, 2H), 2.82 (s, 3H), 2.22 (t, J = 7.3 Hz, 2H), 1.89 – 1.78 (m, 2H), 1.61 – 1.50 (m, 2H), 1.53 (s, 6H), 1.48 – 1.34 (m, 2H).

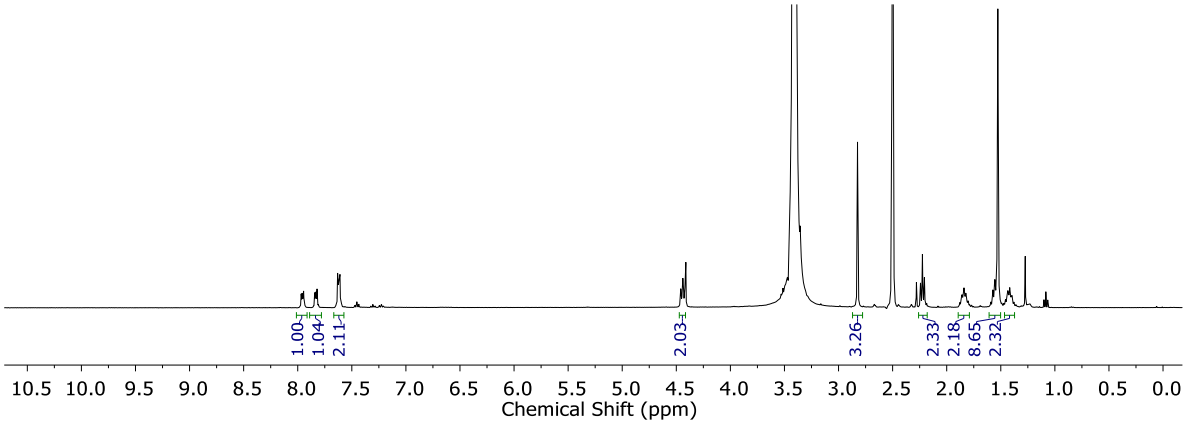
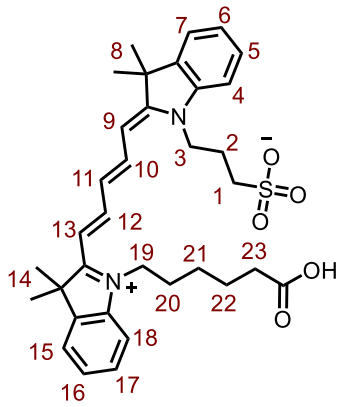


Figure 5.37 | ^{1}H NMR spectrum (400 MHz, 298 K, DMSO-d_6) of **5.K.**

This data was consistent with previously reported data.[10]

3-((Z)-2-((2E,4E)-5-(1-(5-Carboxypentyl)-3,3-dimethyl-3H-indol-1-ium-2-yl)penta-2,4-dien-1-ylidene)-3,3-dimethylindolin-1-yl)propane-1-sulfonate 5.L



Compound **5.K** (200 mg, 0.564 mmol, 1.00 eq.) and malondialdehyde bis(phenylimine) monohydrochloride (175 mg, 0.676 mmol, 1.20 eq.) were combined in Ac_2O (5m L) and heated at 120 °C for 30 min. The solution was cooled to room temperature and a solution of **4.F** (222 mg, 0.789 mmol, 1.40 eq.) in anhydrous pyridine (5 mL) was added. The mixture was stirred at room temperature for 18 h. The solvents were removed under reduced

pressure. The oily residue was purified by silica gel chromatography (CH_3OH in DCM, 0-20%). The solid obtained was dissolved in DCM and precipitated with hexane to give **5.L** as a blue microcrystalline powder (274 mg, 0.464 mmol, 82%). Some impurities (> 5%) remained, which were removed after the following synthetic step.

^{1}H NMR (400 MHz, 298 K, CD_3OD) δ 8.26 (t, *J* = 13.1 Hz, 2H, $H_{10,12}$), 7.48 (dd, *J* = 7.2, 5.5 Hz, 2H, $H_{4,18}$), 7.44 – 7.35 (m, 3H, $H_{5/7/15/17/6/16}$), 7.33 – 7.22 (m, 3H, $H_{5/7/15/17/6/16}$), 6.65 (t, *J* = 12.4 Hz, 1H, H_{11}), 6.43 (d, *J* = 13.7 Hz, 1H, $H_{9/13}$), 6.28 (d, *J* = 13.7 Hz, 1H, $H_{9/13}$), 4.32 (t, *J* = 8.0 Hz, 2H, H_3), 4.11 (t, *J* = 7.6 Hz, 2H, H_{19}), 3.00 (t, *J* = 6.7 Hz, 2H, H_1), 2.32 – 2.19 (m, 4H, $H_{23,2}$), 1.88 – 1.78 (m, 2H, H_{20}), 1.72 (d, *J* = 6.1 Hz, 12H, $H_{8,14}$), 1.75 – 1.67 (m, 2H, H_{22}), 1.57 – 1.44 (m, 2H, H_{21}). **^{13}C NMR** (126 MHz, 298 K, CD_3OD) δ 173.5, 173.0, 154.2, 154.1, 142.1, 141.4, 141.1, 131.6, 128.4, 128.3, 125.7, 124.9, 124.7, 122.0, 122.0, 114.6, 110.7, 110.5, 103.2, 103.1, 49.2, 49.1, 47.7, 43.5, 42.4, 35.2 (b), 26.9, 26.6, 26.5, 26.1, 24.2, 22.6. **HRMS**: *m/z* calculated for [**5.L** + Na^+] = 613.2707, observed = 613.2713.

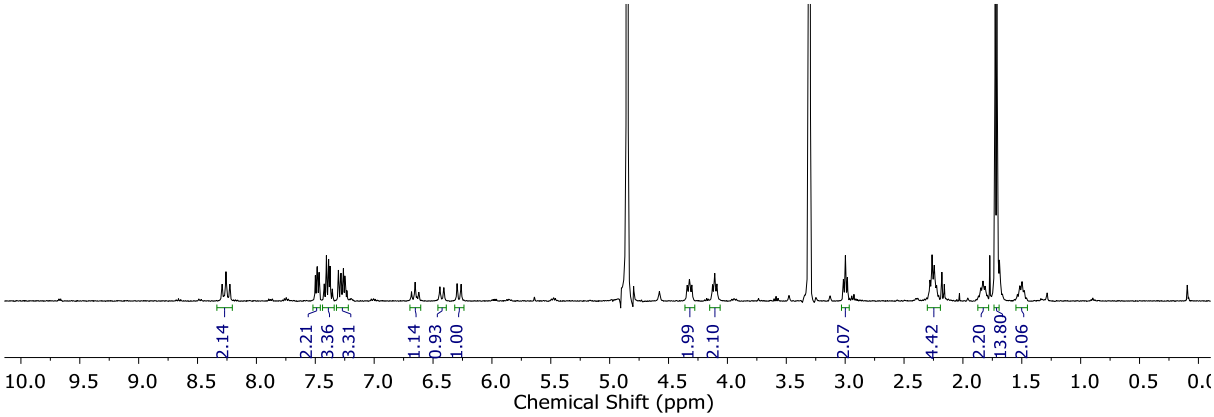


Figure 5.38| ^{1}H NMR spectrum (400 MHz, 298 K, CD_3OD) of **5.L.**

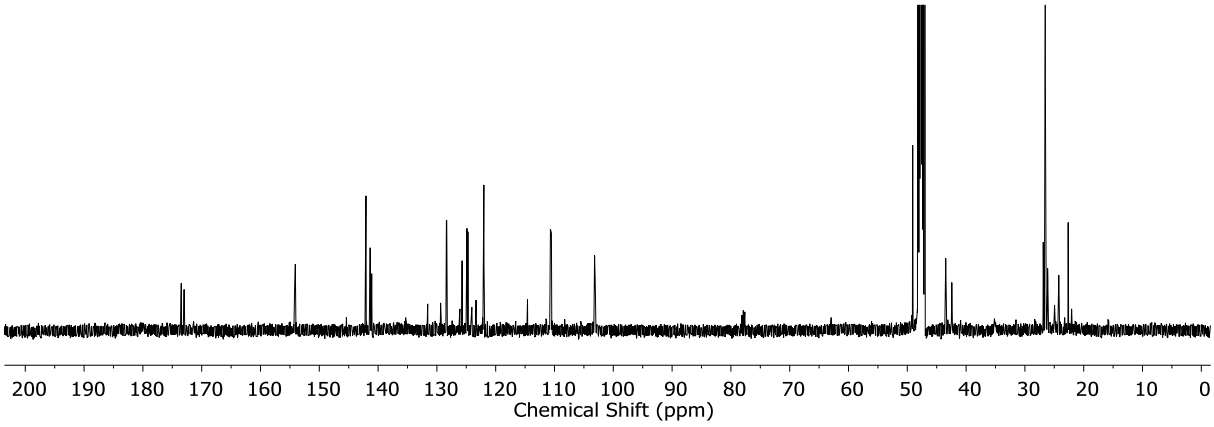


Figure 5.39| ^{13}C NMR (126 MHz, 298 K, CD_3OD) of **5.L**.

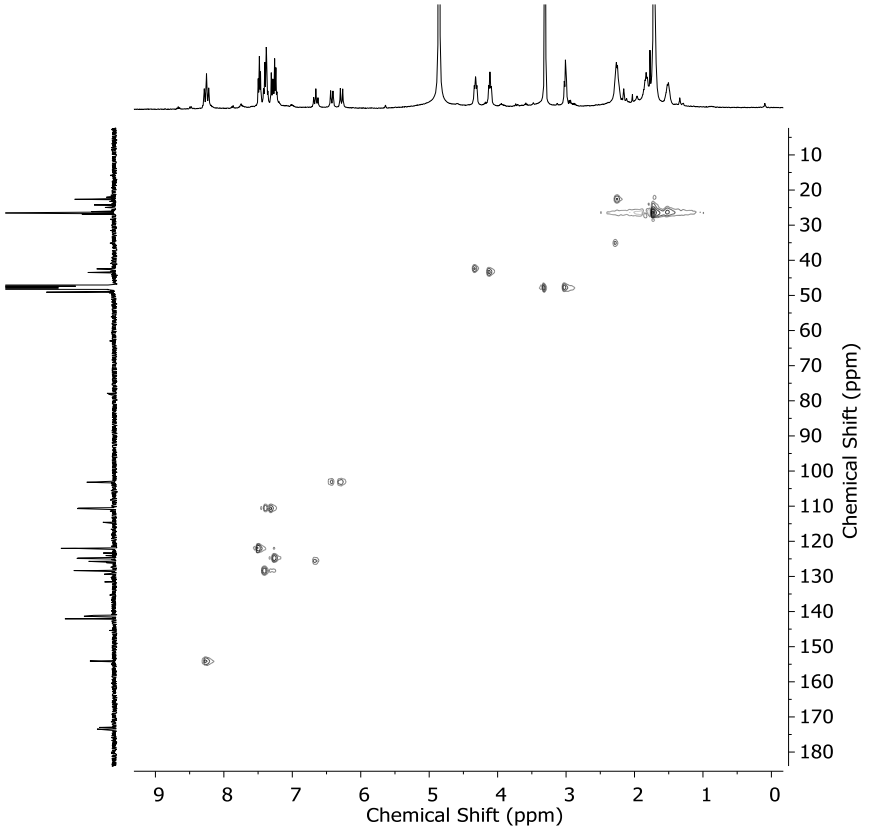


Figure 5.40| ^{1}H-^{13}C HSQC spectrum (400 MHz, 298 K, CD_3OD) of **5.L.**

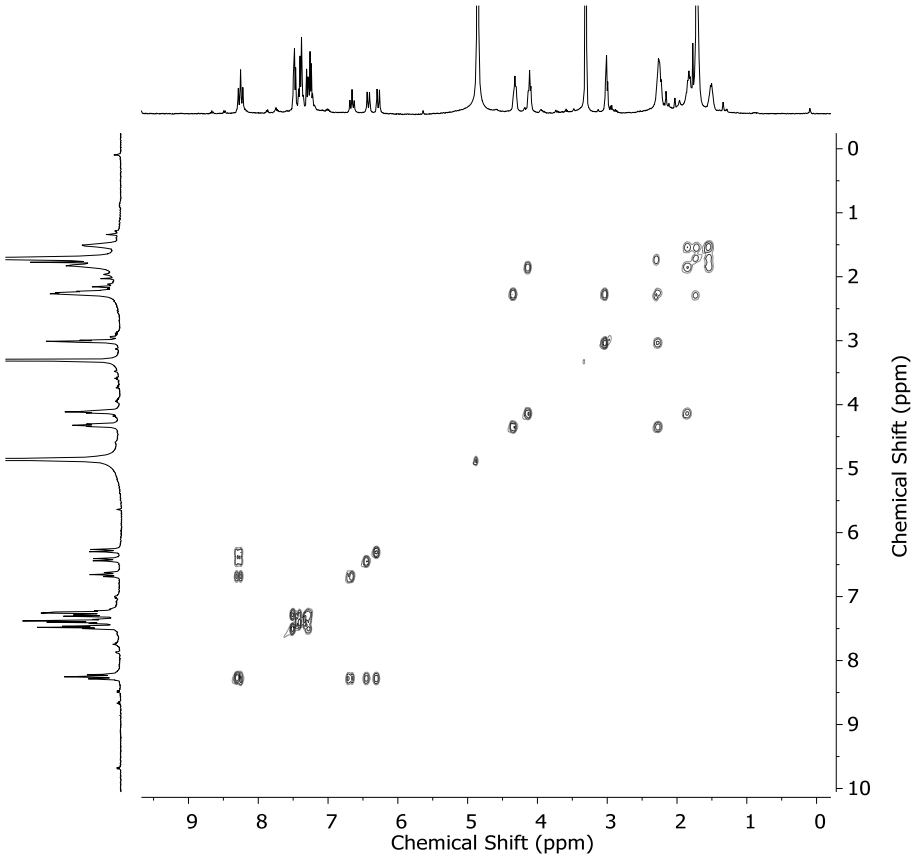
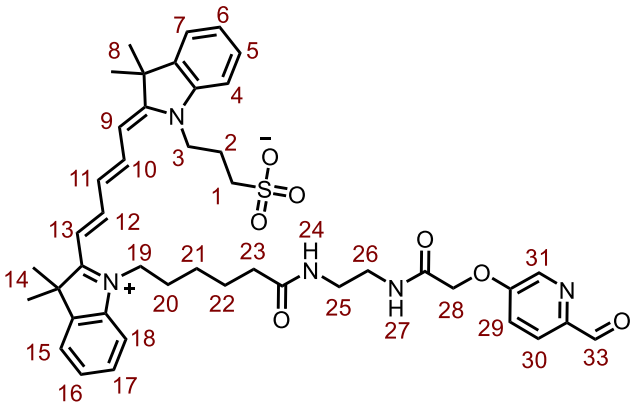


Figure 5.41| ^{1}H-^{1}H COSY spectrum (400 MHz, 298 K, CD_3OD) of **5.L.**

3-((Z)-2-((2E,4E)-5-(1-(6-((2-(2-((6-Formylpyridin-3-yl)oxy)acetamido)ethyl)amino)-6-oxohexyl)-3,3-dimethyl-3H-indol-1-ium-2-yl)penta-2,4-dien-1-ylidene)-3,3-dimethylindolin-1-yl)propane-1-sulfonate 5.M



Compound **5.L** (50 mg, 0.085 mmol, 1.00 eq.), compound **5.I** (25 mg, 0.085 mmol, 1.00 eq.), HBTU (33 mg, 0.087 mmol, 1.02 eq.) and DIPEA (117 uL, 0.671 mmol, 7.90 eq.) were combined in DMF (1.5 mL) and stirred at room temperature for 2 h. The product was precipitated with Et_2O, dissolved in DCM (1.0 mL) and precipitated with Et_2O. The solid was washed with EtOAc and Et_2O. The solid was dried under a flow of N_2 to give **5.M** as a dark blue solid (62 mg, 0.078 mmol, 91%).

^{1}H NMR (500 MHz, 298 K, DMSO-d_6) δ 9.84 (s, 1H, H_{32}), 8.48 (d, J = 2.7 Hz, 1H, H_{31}), 8.35 – 8.28 (m, 2H, $H_{10,12}$), 8.28 (t, J = 5.2 Hz, 1H, H_{27}), 7.90 (d, J = 8.7 Hz, 1H, H_{30}), 7.88 (t, J = 5.6 Hz, 1H, H_{24}), 7.61 – 7.55 (m, 2H, $H_{4,18}$), 7.52 (dd, J = 8.7 Hz, J = 2.7 Hz, 1H, H_{29}), 7.47 – 7.31 (m, 4H, $H_{5,7,15,17}$), 7.26 – 7.17 (m, 2H, $H_{6,16}$), 6.53 (t, J = 12.3 Hz, 1H, H_{11}), 6.42 (d, J = 13.7 Hz, 1H, $H_{9/13}$), 6.27 (d, J = 13.7 Hz, 1H, $H_{9/13}$), 4.69 (s, 2H, H_{28}), 4.25 (t, J = 7.8 Hz, 2H, H_3), 4.04 (t, J = 7.4 Hz, 2H, H_{19}), 3.13 (2t, J = 5.9 Hz, 4H, $H_{26,25}$), 2.57 (t, J = 6.8 Hz, 2H, H_1), 2.07 – 1.93 (m, 4H, $H_{23,2}$), 1.65 (d, J = 3.8 Hz, 14H, $H_{14,8,20}$), 1.54 – 1.46 (m, 2H, H_{22}), 1.36 – 1.27 (m, 2H, H_{21}). **^{13}C NMR** (126 MHz, 298 K, DMSO-d_6) δ 192.1, *172.9*, 172.4, *172.7*, 166.9, *161.0*, 157.4, *154.3*, *154.1*, *151.9*, *150.4*, 146.1, 142.1, 141.2, 139.1, *128.6*, *125.7*, *124.8*, 123.6, 122.5, 121.5, 111.2, 111.1, *103.5*, *103.2*, 67.1, 65.0, 59.8, 49.0, 48.9, 47.9, *43.74*, *42.8*, 38.6, 38.1, 35.2, 27.3, 27.2, 26.8, 25.8, 24.9, 23.4. The values in italics were found by correlation in HSQC and HMBC spectra. **HRMS**: *m/z* calculated for [**5.M** + Na^+] = 818.3558, observed = 818.3569.

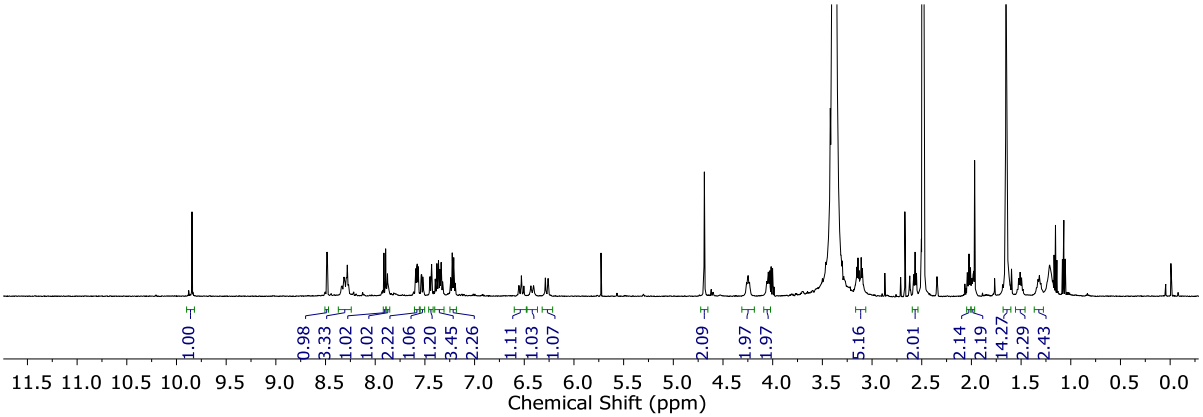


Figure 5.42| ^{1}H NMR spectrum (500 MHz, 298 K, DMSO-d_6) of **5.M.**

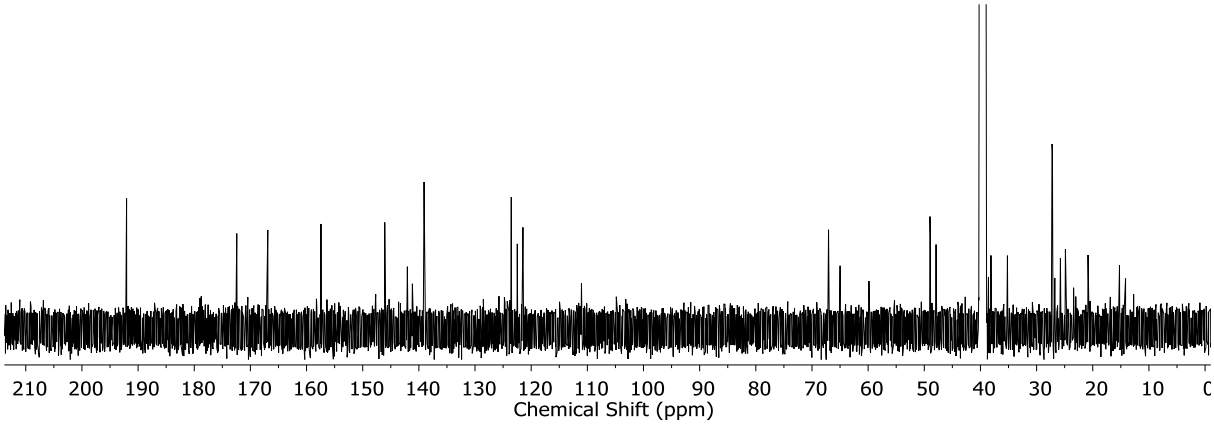


Figure 5.43| ^{13}C NMR (126 MHz, 298 K, DMSO-d_6) of **5.M**.

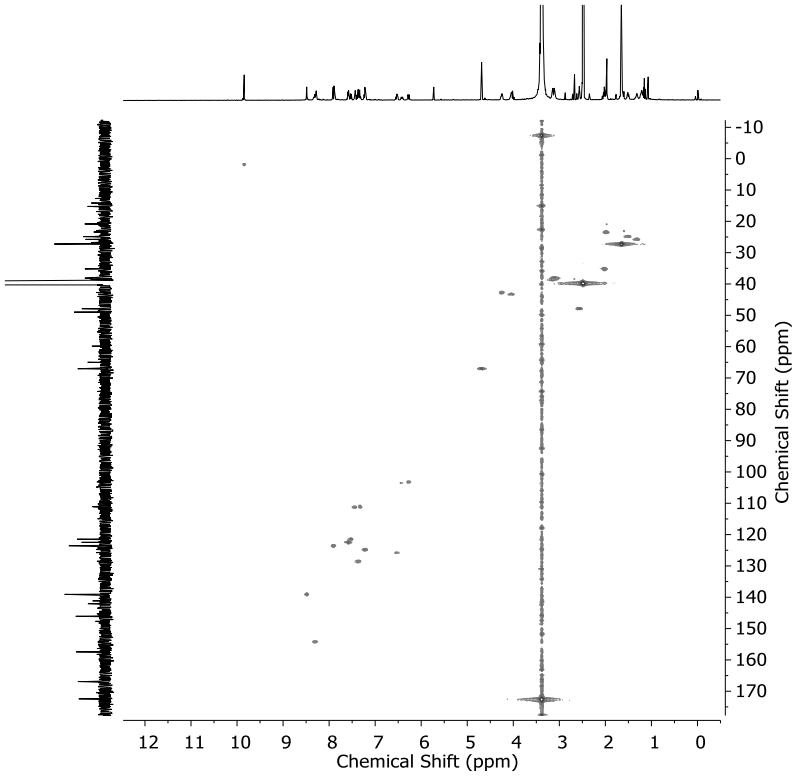


Figure 5.44| ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, DMSO-d_6) of **5.M.** Spectral aliasing for the signal corresponding to the aldehyde ^{13}C resonance (actual shift = 192 ppm).

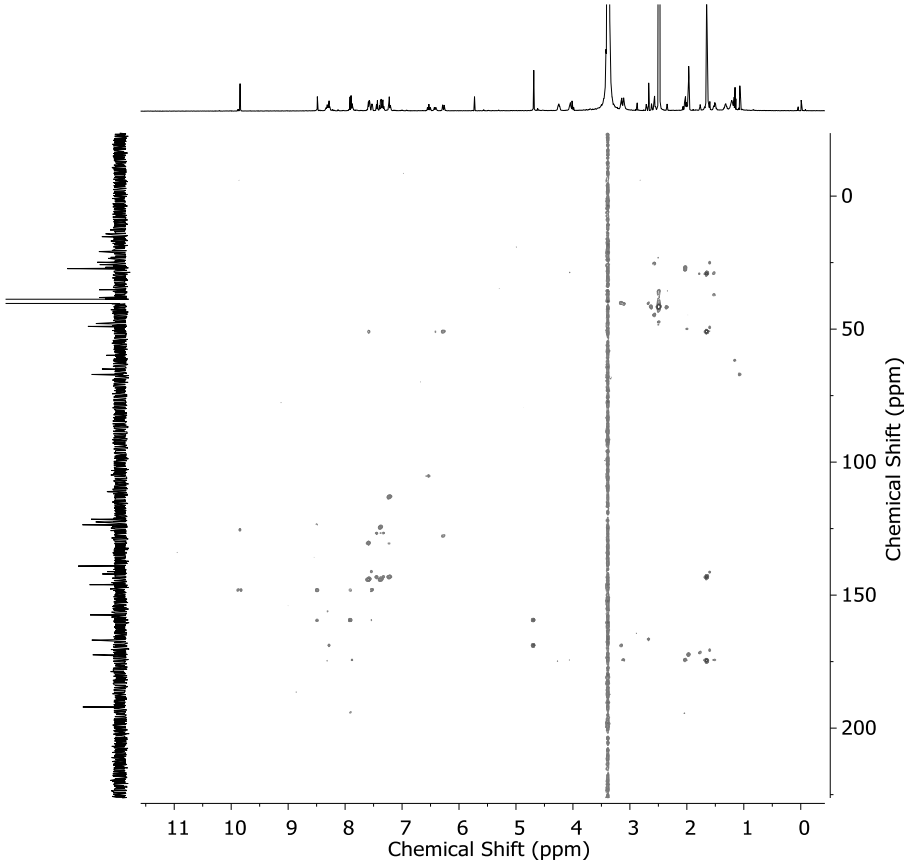


Figure 5.45| ^{1}H-^{13}C HMBC spectrum (500 MHz, 298 K, DMSO-d_6) of **5.M.**

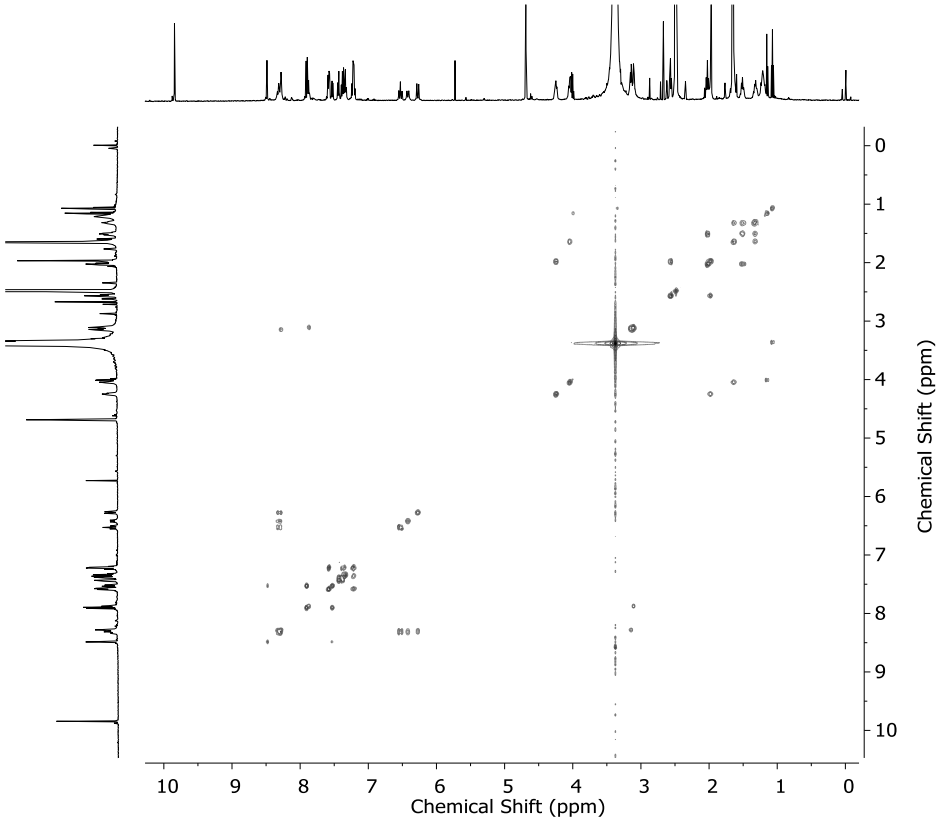
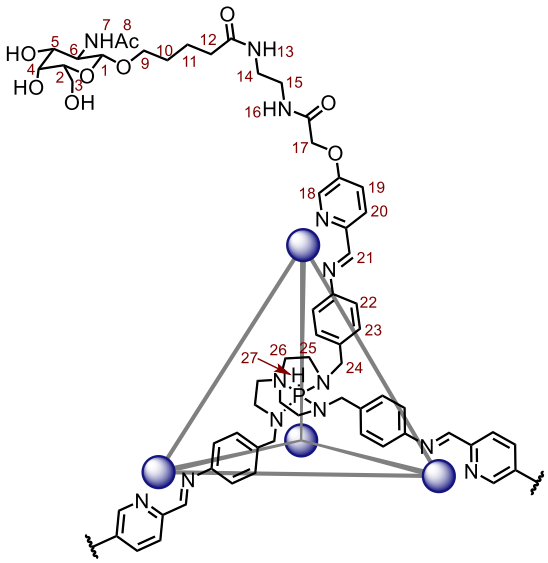


Figure 5.46| ^{1}H-^{1}H COSY spectrum (500 MHz, 298 K, DMSO-d_6) of **5.J.**

5.6.3 Synthesis of supramolecular cages

Synthesis of cage 5.1



Compound **5.J** (6.5 mg, 12.2 µmol, 12.0 eq.), compound **5.N** (2.2 mg, 4.1 µmol, 4.0 eq.), $FeSO_4$ (1.1 mg, 4.1 µmol, 4.0 eq.) and $NaReO_4$ (0.3 mg, 1.0 µM, 1.0 eq.) were combined in D_2O (0.5 mL) and stirred at room temperature for 18 h. The sample was used without further purification.

^{1}H NMR (500 MHz, 298 K, D_2O) δ 8.63 (s, 12H, H_{21}), 8.40 (d, J = 8.8 Hz, 12H, H_{20}), 7.79 (d, J = 8.8 Hz, 12H, H_{19}), 7.12 – 6.91 (m, 36H, $H_{18,23}$), 6.19 (d, J = 497 Hz, 4H, H_{27}), 4.25 (d, J = 8.5 Hz, 12H, H_1), 3.81 – 3.31 (m, 120H, $H_{2\text{-}6,9,24}$), 3.24 – 3.09 (m, 48H, $H_{14\text{-}15}$), 2.71 – 2.40 (m, 48H, $H_{25\text{-}26}$), 1.83 (s, 36H, H_8), 1.55 – 1.24 (m, 48H, $H_{10\text{-}11}$). H_{17} and H_{22} were found under the water peak and H_{12} under the residual acetone peak. **^{13}C NMR** (126 MHz, 298 K, D_2O) δ 177.2, 174.6, 172.6, 169.2, 157.4, 151.5, 149.3, 145.4, 136.6, 131.8, 130.1, 122.5, 101.5, 75.0, 70.9, 69.7, 67.7, 66.8, 60.9, 54.3, 52.3, 42.5, 38.6, 37.8, 35.3, 33.1, 28.0, 22.2, 21.9. The signal for C_{22} was assumed too broad to be detectable in the ^{13}C NMR spectrum.

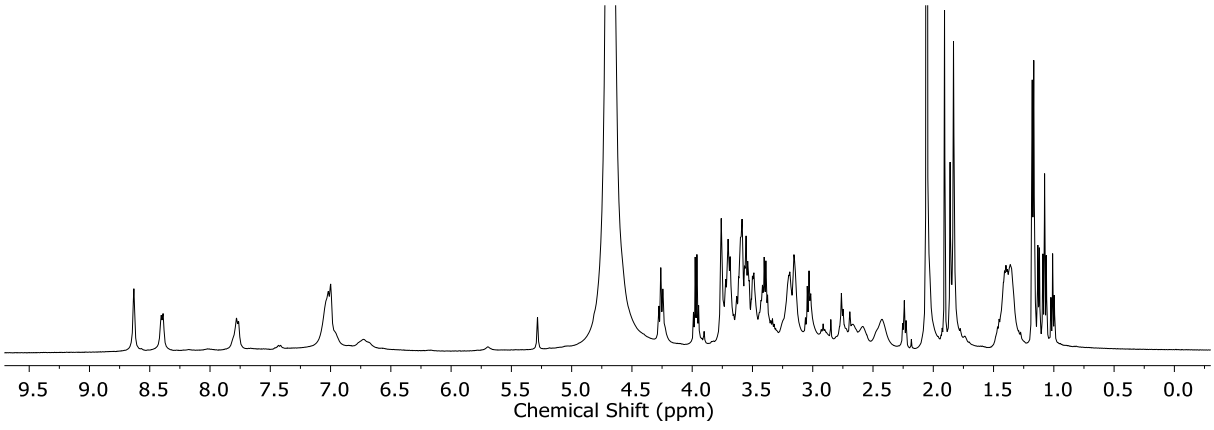


Figure 5.47 | ^{1}H NMR spectrum (500 MHz, 298 K, D_2O) of **5.1.**

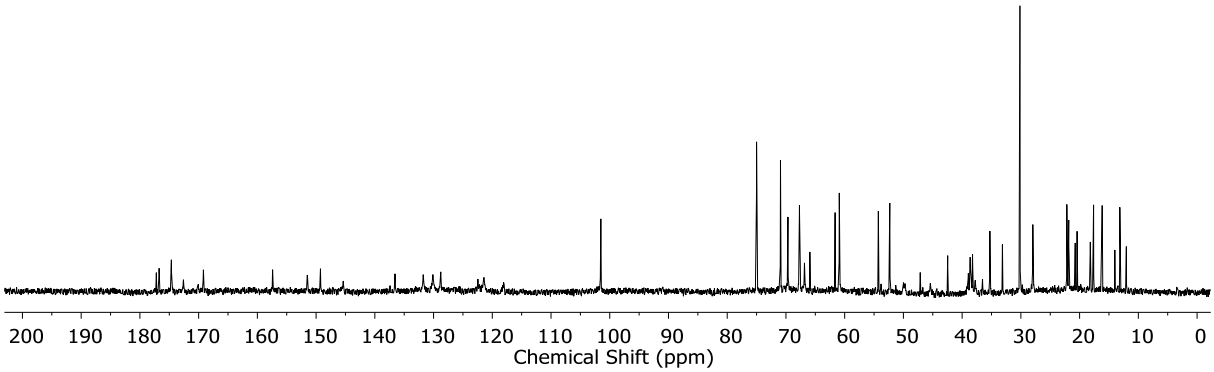


Figure 5.48 | ^{13}C NMR (126 MHz, 298 K, D_2O) of **5.1.**

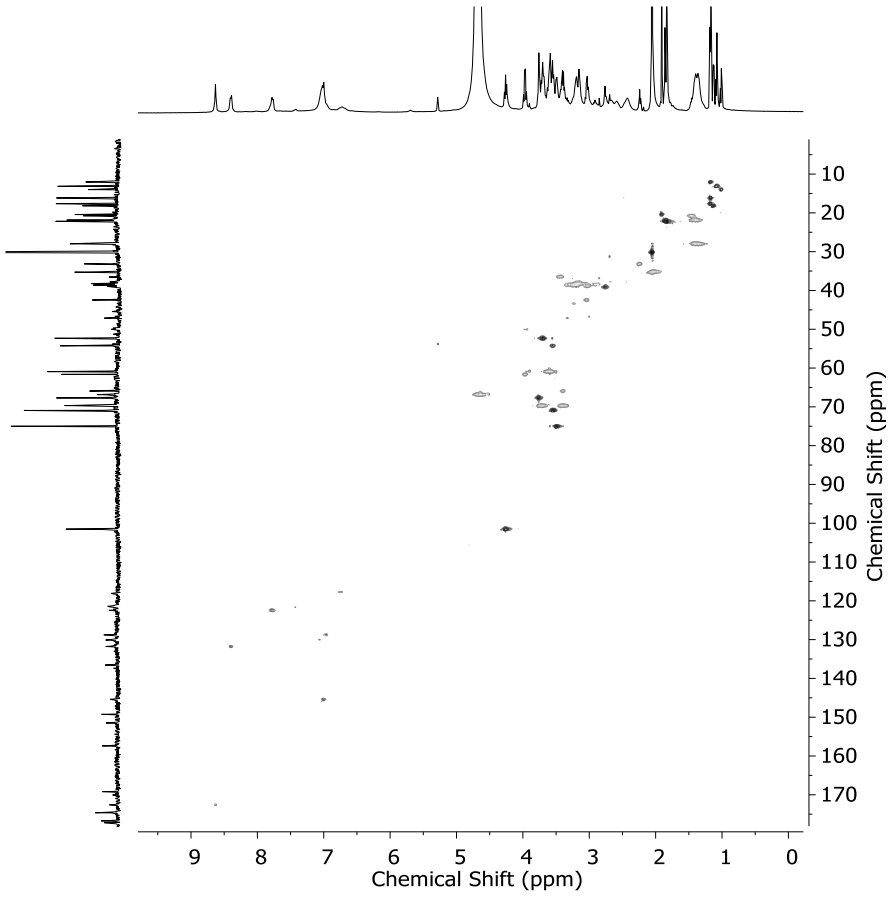


Figure 5.49 | ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, D_2O) of **5.1**

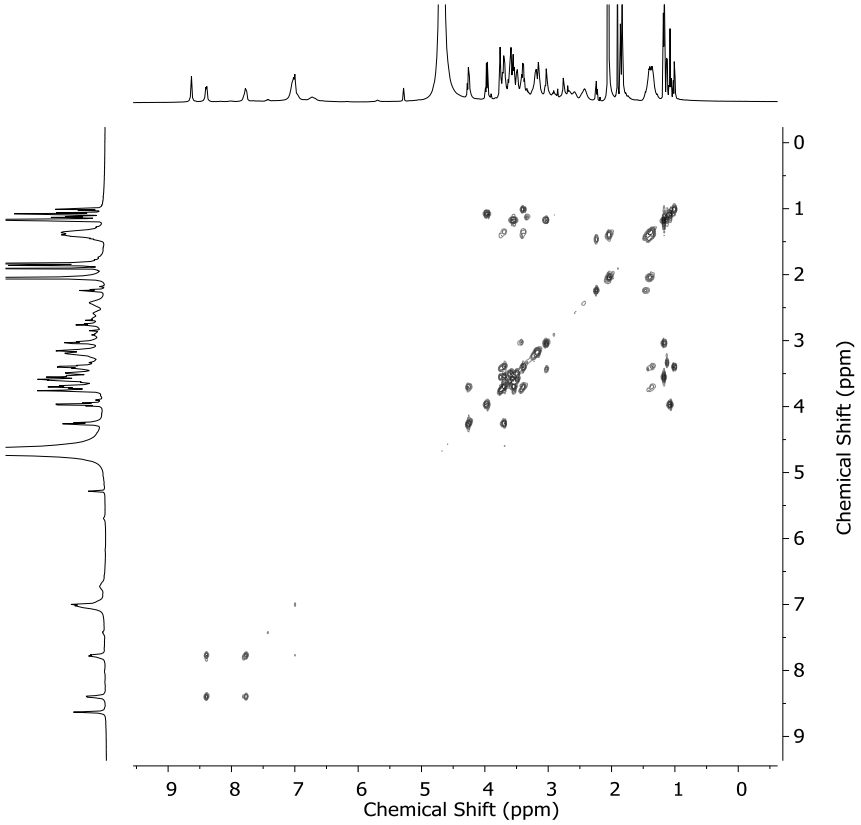
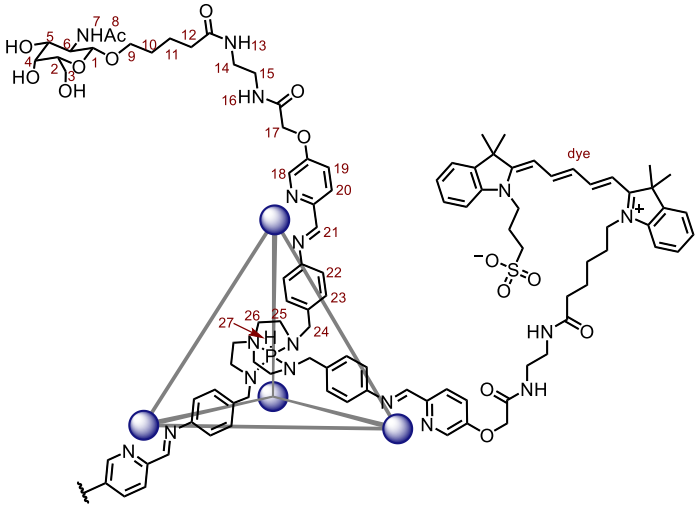


Figure 5.50 | ^{1}H-^{1}H COSY spectrum (500 MHz, 298 K, D_2O) of **5.1.**

Synthesis of cage 5.2



Compound **5.J** (8.2 mg, 15.6 µmol, 10.0 eq.), compound **5.N** (3.3 mg, 6.3 µmol, 4.0 eq.), $FeSO_4$ (1.7 mg, 6.3 µM, 4.0 eq.) and $NaReO_4$ (0.4 mg, 1.3 µM, 1.0 eq.) were combined in D_2O (0.5 mL). Compound **5.M** (2.5 mg, 3.1 µmol, 2.0 eq.) was dissolved in CH_3CN (0.1 mL). The two solutions were combined and stirred at room temperature for 18 h. The CH_3CN was evaporated under a flow of N_2 and the sample was filtered on glass fibre before being used.

^{1}H NMR (500 MHz, 298 K, D_2O) δ 8.63 (bs, 12H, H_{21}), 8.39 (bd, J = 8.8 Hz, 12H, H_{20}), 7.95 (d, J = 8.3 Hz, xH, H_{dye}), 7.86 (d, J = 8.2 Hz, xH, H_{dye}), 7.76 (bd, J = 8.8 Hz, 12H, H_{19}), 7.65 (d, J = 8.3 Hz, xH, H_{dye}), 7.56 (d, J = 8.7 Hz, xH, H_{dye}), 7.12 – 6.91 (m, 36H, $H_{18,23}$), 6.19 (d,

J = 497 Hz, 4H, H_{27}), 4.26 (d, *J* = 8.7 Hz, 12(1-x)H, H_1), 3.80 – 3.34 (m, 120(1-x)H, $H_{2\text{-}6,9,24}$), 3.26 – 3.09 (m, 48(1-x)H, $H_{14\text{-}15}$), 2.67 – 2.35 (m, 48H, $H_{25\text{-}26}$), 2.04 (s, 24(1-x)H, H_{12}), 1.83 (s, 36(1-x)H, H_8), 1.52 – 1.25 (m, 48(1-x)H, $H_{10\text{-}11}$). H_{17} and H_{22} were found under the water peak. **^{13}C NMR** (126 MHz, 298 K, D_2O) δ 177.2, 174.6, 172.6, 169.2, 157.4, 151.5, 149.3, 145.4, 136.6, 131.8, 130.1, 122.5, 101.5, 75.0, 70.9, 69.7, 67.7, 66.8, 60.9, 54.3, 52.3, 42.5, 38.6, 37.8, 35.3, 33.1, 28.0, 22.2, 21.9. The signal for C_{22} was assumed too broad to be detectable in the ^{13}C NMR spectrum. The intensity of the signals for the dye were too low to be observed, except for a few aromatic protons.

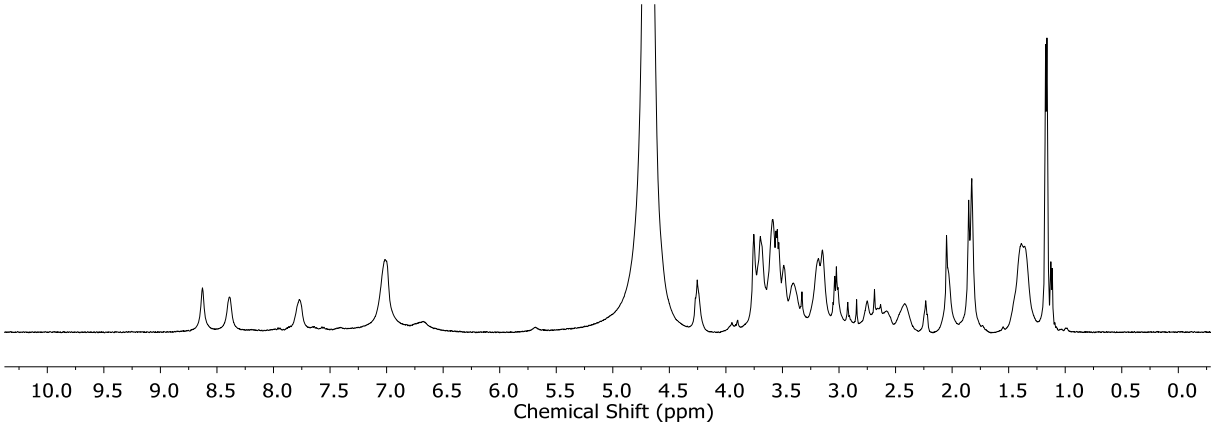


Figure 5.51 | ^{1}H NMR spectrum (500 MHz, 298 K, D_2O) of **5.2.**

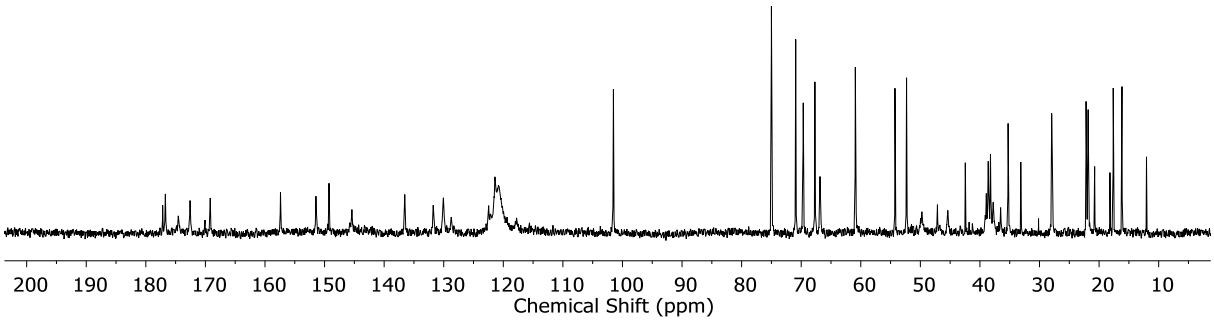


Figure 5.52 | ^{13}C NMR (126 MHz, 298 K, D_2O) of **5.2.**

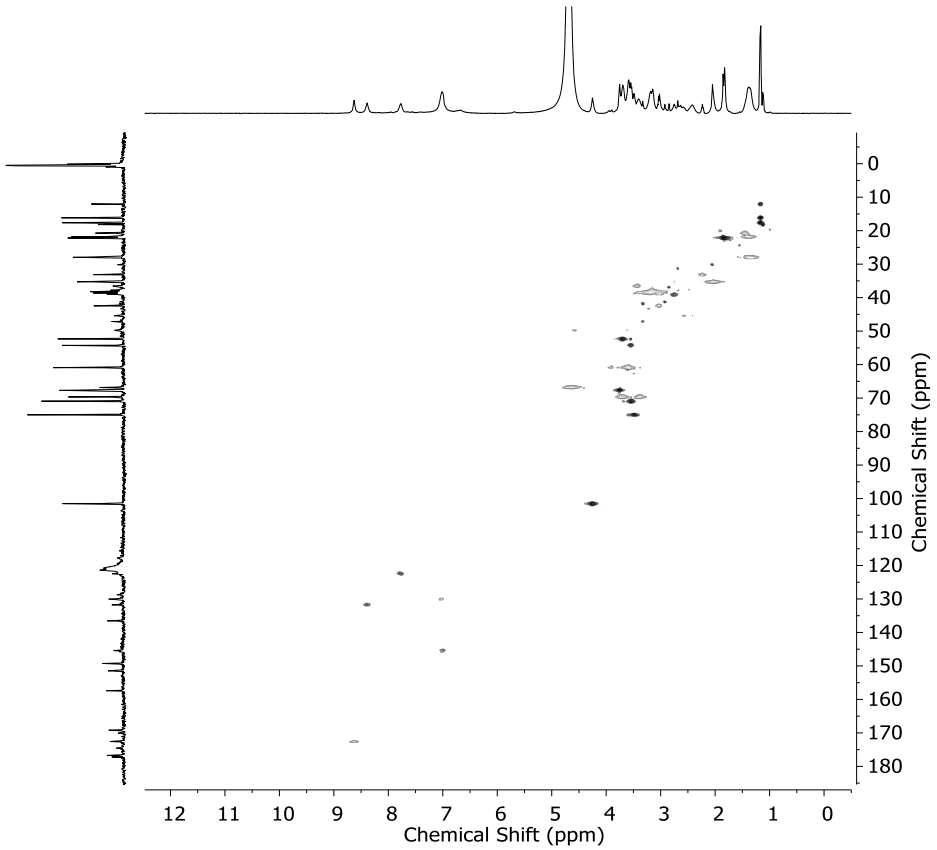


Figure 5.53| ^{1}H-^{13}C HSQC spectrum (500 MHz, 298 K, D_2O) of **5.2.**

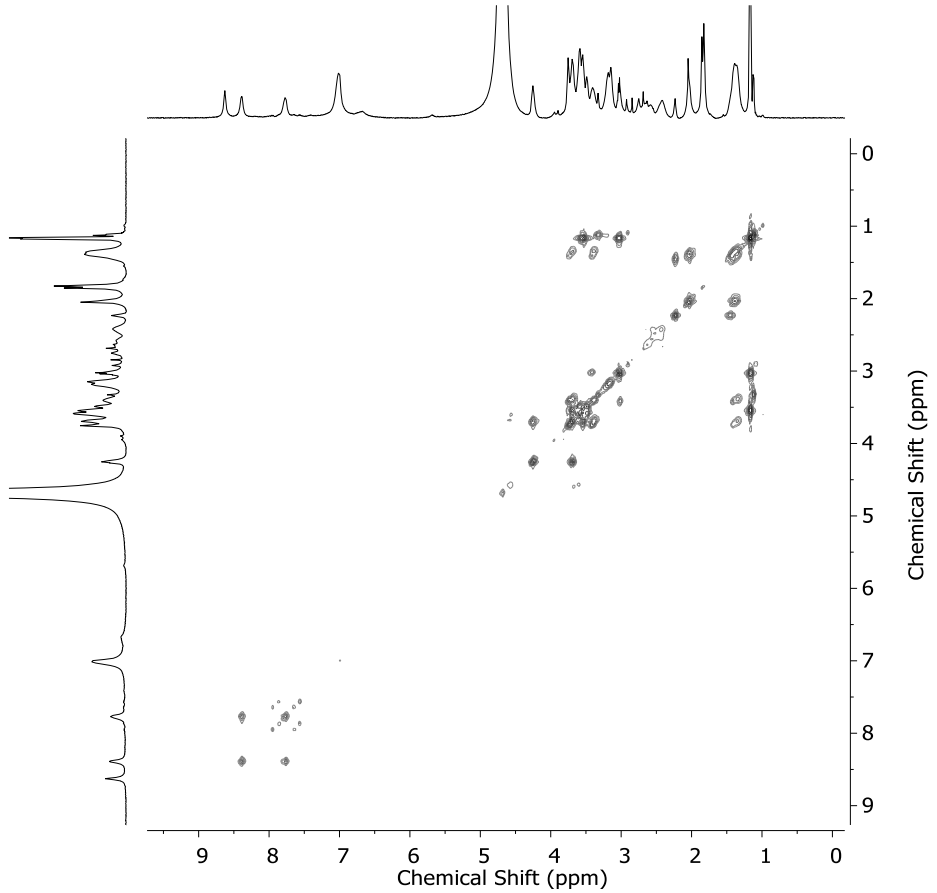


Figure 5.54| ^{1}H-^{1}H COSY spectrum (500 MHz, 298 K, D_2O) of **5.2.**

5.7 References

[1] J. Han, A. F. B. Räder, F. Reichart, B. Aikman, M. N. Wenzel, B. Woods, M. Weinmüller, B. S. Ludwig, S. Stürup, G. M. M. Groothuis, H. P. Permentier, R. Bischoff, H. Kessler, P. Horvatovich, A. Casini, *Bioconjugate Chem.* **2018**, *29*, 3856-3865.

[2] M. D. Patel, V. C. Malshe, V. Pujari, A. Gorakshakar, M. Madkaikar, K. Ghosh, P. V. Devarajan, *Drug Deliv.* **2017**, *24*, 20-29.

[3] A. A. D'Souza, P. V. Devarajan, *J. Controlled Release* **2015**, *203*, 126-139.

[4] X. Huang, J.-C. Leroux, B. Castagner, *Bioconjugate Chem.* **2017**, *28*, 283-295.

[5] C. Bradshaw, W. , S. Sakamuri, D. Liu, *Solstice Biologics, LTD.* **2016**, WO/2016/094677.

[6] B. Rajwa, T. Bernas, H. Acker, J. Dobrucki, J. P. Robinson, *Microsc. Res. Tech.* **2007**, *70*, 869-879.

[7] L. D. Lavis, R. T. Raines, *ACS Chem. Biol.* **2008**, *3*, 142-155.

[8] *ThermoFisher Scientific*, https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/fluorophores/alexa-fluor-647.html.

[9] E. Herbst, D. Shabat, *Org. Biomol. Chem* **2016**, *14*, 3715-3728.

[10] M. V. Kvach, A. V. Ustinov, I. A. Stepanova, A. D. Malakhov, M. V. Skorobogatyi, V. V. Shmanai, V. A. Korshun, *Eur. J. Org. Chem.* **2008**, *2008*, 2107-2117.

[11] D. Zhang, T. K. Ronson, J. Mosquera, A. Martinez, L. Guy, J. R. Nitschke, *J. Am. Chem. Soc.* **2017**, *139*, 6574-6577.

[12] E. G. Percástegui, J. Mosquera, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2017**, *56*, 9136-9140.

[13] B. P. Burke, W. Grantham, M. J. Burke, G. S. Nichol, D. Roberts, I. Renard, R. Hargreaves, C. Cawthorne, S. J. Archibald, P. J. Lusby, *J. Am. Chem. Soc.* **2018**, *140*, 16877-16881.

[14] P. R. Symmers, M. J. Burke, D. P. August, P. I. T. Thomson, G. S. Nichol, M. R. Warren, C. J. Campbell, P. J. Lusby, *Chem. Sci.* **2015**, *6*, 756-760.

[15] M. Dowlut, D. G. Hall, O. Hindsgaul, *J. Org. Chem.* **2005**, *70*, 9809-9813.

[16] H. Lemoine, D. Marković, B. Deguin, *J. Org. Chem.* **2014**, *79*, 4358-4366.

[17] D. J. Wilson, C. Shi, B. P. Duckworth, J. M. Muretta, U. Manjunatha, Y. Y. Sham, D. D. Thomas, C. C. Aldrich, *Anal. Biochem.* **2011**, *416*, 27-38.

[18] T. P. Prakash, M. J. Graham, J. Yu, R. Carty, A. Low, A. Chappell, K. Schmidt, C. Zhao, M. Aghajan, H. F. Murray, S. Riney, S. L. Booten, S. F. Murray, H. Gaus, J. Crosby, W. F. Lima, S. Guo, B. P. Monia, E. E. Swayze, P. P. Seth, *Nucleic Acids Res.* **2014**, *42*, 8796-8807.

[19] T. M. Wickramaratne, V. C. Pierre, *Bioconjugate Chem.* **2015**, *26*, 63-70.

Chapter 6

Species quantification in dynamic libraries of self-assembled coordination cages by mass spectrometry – Three case-studies

6.1 Introduction

The molecular components of biological systems self-sort in different ways to function synergically and to avoid interfering with each other. Similarly, in synthetic supramolecular coordination cages, molecules can interact in different ways to yield complex self-assembled libraries.[1-7] The self-assembly of multiple components can result in the clean formation of structurally complex single products,[8-13] as each building block is guided to a specific location during thermodynamic equilibration. Understanding the behaviour of dynamic mixtures and the driving forces leading to some combinations of components forming discrete entities,[14] while others do not is crucial for the future elaboration of more complex systems. In this process, rules governing self-assembly are often uncovered[15-16] allowing for the design of increasingly complex self-assembling synthetic systems and materials.[17]

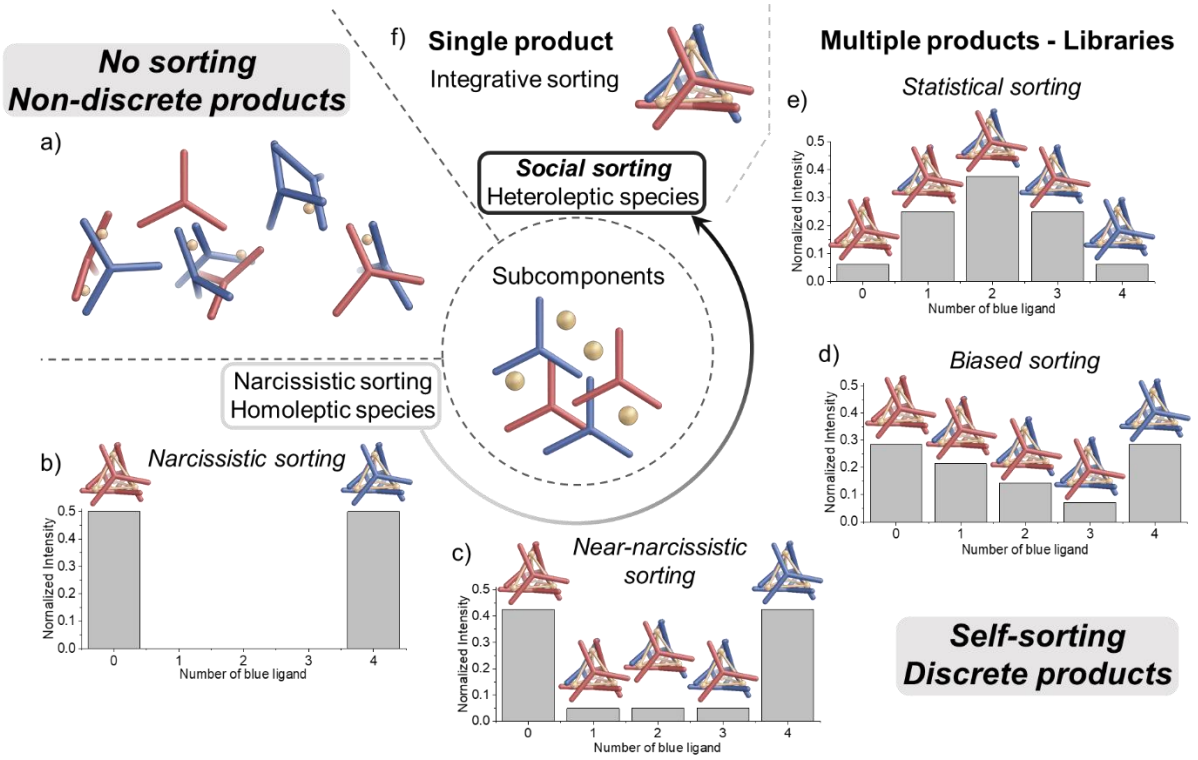


Figure 6.1 | Representation of the different outcomes when self-assembly is carried using different ligands where either a) no sorting, or b and f) self-sorting giving discrete products occurs. Within self-sorting f) a single integrative product or b and e) libraries of products can be formed, which range from homoleptic species only (b – narcissistic sorting) to various types of social sorting giving both homo- and heteroleptic species (c – near-narcissistic, d – biased or e – statistical sorting).

In many cases, however, mixtures of different subcomponents produce multiple self-assembled products.[18-20] Libraries of structures are formed which generally fall within one of the two categories: social self-sorting (Figure 6.1 c-f) or narcissistic self-sorting (Figure 6.1 b). The

latter is defined by a complete segregation of the subcomponents in order to form exclusively the homoleptic complexes. Within social sorting, where ligands combine to form heteroleptic complexes preferentially, different regimes can be adopted. If no energetic bias for any product exist, the products will distribute statistically (following the binomial law for two subcomponents for example, Figure 6.1 e). When an enthalpic bias is present, the formation of a subgroup of all the possible species will be favoured (Figure 6.1 d). For example, if a type of geometry is formed preferentially, the sorting can be referred to as geometrically biased sorting. The formation of the least favoured species, in reduced proportions, also occurs in these cases.

Deciphering the self-assembly rules within such systems involves new characterization challenges: interconverting, low-symmetry, and sometimes paramagnetic products in the case of coordination complexes are difficult to detect by NMR as they lead to multiple broad, overlapping signals. Labile species can re-equilibrate after chromatographic separation due to their dynamic behaviour.[21] Mass spectrometry could potentially overcome these challenges by giving an instantaneous picture of the system, allowing the speciation of components within complex mixtures to be gauged.

This chapter presents, through three case-studies, some of the different types of sorting regimes which can be encountered in self-assembly and the potential to analyse mixtures of self-sorted assemblies and assess individual outcomes by mass spectrometric techniques. In the first system, the relative energies of heteroleptic structures compared to the more stable homoleptic were obtained, allowing for the quantification of each ligand's structural preferences.[22] In a second case, quantitative information on both the product distribution, and the proportion of ligands incorporated into each species within mixtures was obtained for a complex system of self-assembled scalenohedra or pseudo-octahedra using a unique combination of ESI-MS and IM-MS.[23] Following a similar approach to the first case, the effect of anion encapsulation on a dynamic library of self-assembled tetrahedra was analysed.

6.2 Energetic quantification of structural preference between tetrahedra and tetragonal prisms

Elucidation of the structural preferences that favour the formation of *mer* vertices over *fac* would allow the directed formation of larger structures such as prisms[24-26] as opposed to the most commonly observed tetrahedra.[27-30] With increased knowledge of the rules behind self-assembly, better control over the rational design of ligands could therefore be achieved.

6.1.1 Chemical system

Van der Waals surfaces of MM3-optimised models of *fac* vertices were calculated for pyridylimines coordinated around octahedral metal centres (Figure 6.2 a and b). The surfaces showed a sterically crowded environment around the aromatic substituent at the 5-position on the pyridyl rings. Electrostatic potential maps of *mer* vertices were also obtained, showing greater potential for π-stacking interactions between the electron-rich aniline rings and the electron deficient pyridyl rings for *mer* vertices (Figure 6.2 c). It was therefore postulated that, by making the aromatic substituent at the 5-position more sterically demanding and more electron withdrawing, the formation of *fac* vertices might be disfavoured relative to *mer*. Using a 2,3,5,6-tetrafluorobenzene ring appended to the pyridyl ring was hypothesised to fulfil both of these requirements due to the larger van der Waals radius of fluorine (1.47 Å) compared to hydrogen (1.00 Å), and the electron-withdrawing nature of the fluorine atoms.[31-32]

Two pyridine based ligands, 2-formyl-5-phenylpyridine **6.A** or 2-formyl-5-(2,3,5,6-tetrafluorophenyl)pyridine **6.B**, presenting either a phenyl or a perfluorophenyl ring at the 5-position of the pyridine were synthesised. Upon self-assembly with *p*-fluoroaniline and $Fe(NTf_2)_2$ in acetonitrile, mononuclear complexes **6.1** and **6.2** were formed respectively (Figure 6.3). ESI-MS results were consistent with the formation of mononuclear complexes of $Fe^{II}L^{A}_3$ or $Fe^{II}L^{B}_3$ stoichiometry (where L^{A} and L^{B} represent ligands incorporating subcomponents **6.A** and **6.B**, respectively). The ^{1}H NMR spectrum of **6.1** indicated the presence of both *fac* and *mer* geometries in the statistical proportion of 1:3. Unlike **6.1**, the ^{1}H NMR spectrum of **6.2** indicated the presence of *mer* geometry exclusively (Figure 6.3). These results indicated that **6.A** can assemble into complexes of both geometries and as a consequence does not show a clear preference for either *fac* or *mer* coordination. Strikingly, **6.B** showed a clear preference to undergo *mer* coordination exclusively.

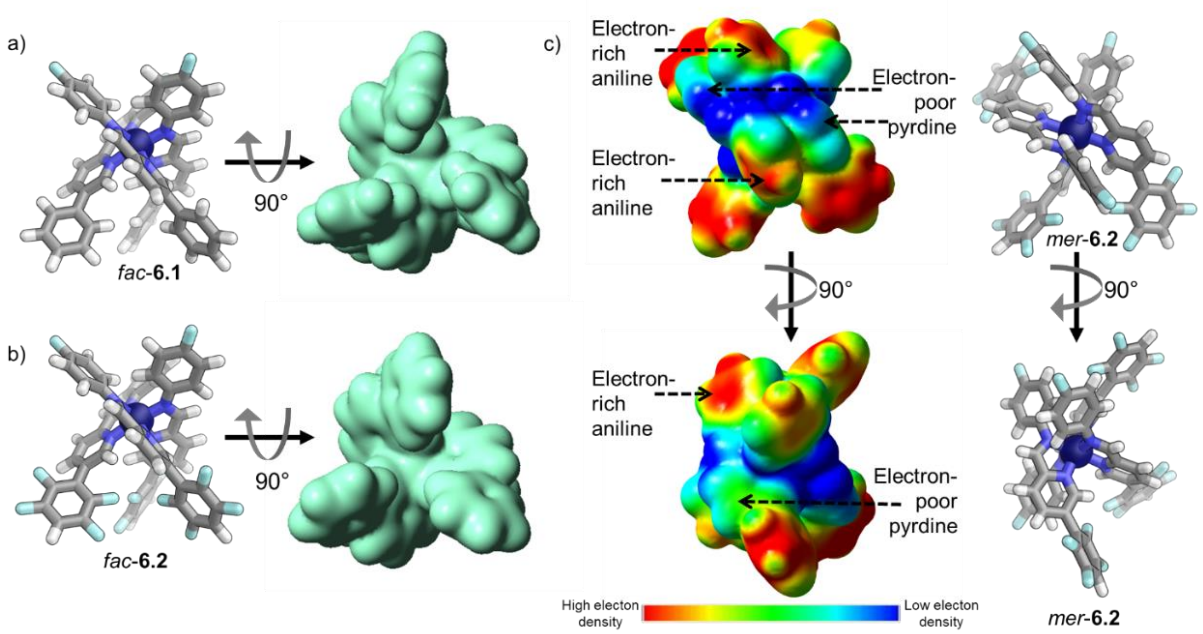


Figure 6.2 | a and b) Van der Waals surfaces (green) calculated from optimised geometries by energy minimization using the MO-G PM6 method (left) on the software package SCIGRESS 3.1.9 for *fac* vertices with (**6.2**) or without (**6.1**) fluorine atoms. c) Electrostatic Potential Map (EPM) of *mer*-**6.2** calculated from the geometry optimised by energy minimization using the PM6-D3H4 method with AMPAC 10.1.[33]

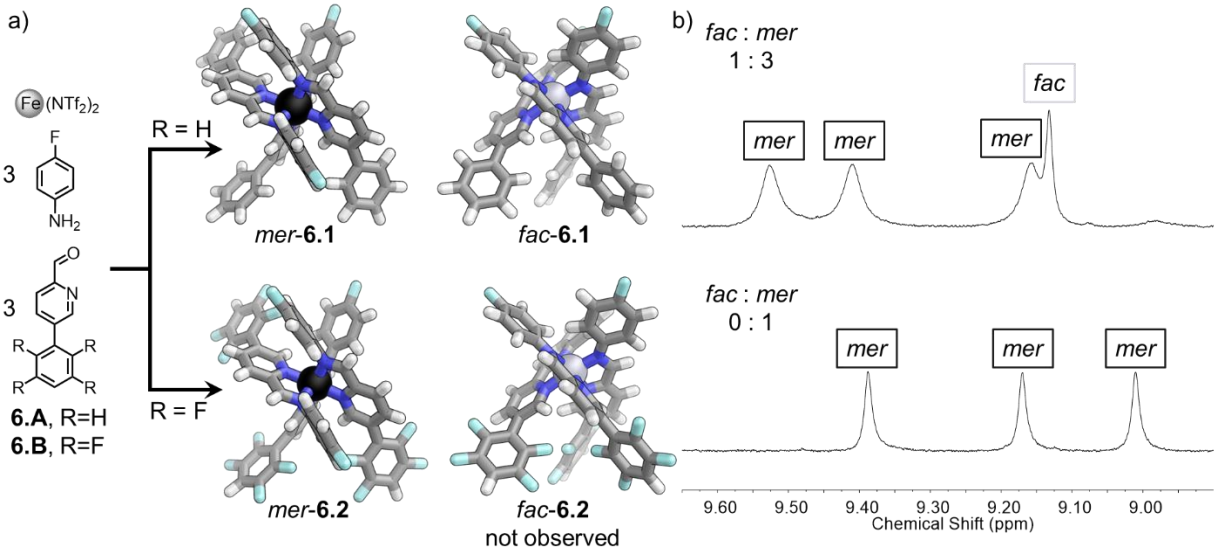


Figure 6.3 | a) Self-assembly of *fac* and *mer* mononuclear complexes **6.1** and **6.2**. b) ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) of the imine regions of **6.1** and **6.2**.

Dialdehyde subcomponent **6.C** is known to self-assemble with *p*-methoxyaniline **6.E** and $Fe(PF_6)_2$ to form $Fe^{II}{}_4L^C{}_6$ tetrahedral cage **6.3** in solution, where L^C represent the ligand incorporating subcomponent **6.C** (Figure 6.4).[34] A similar dialdehyde subcomponent to **6.C**, which contains a 2,3,5,6-tetrafluorobenzene moiety bridging the two formylpyridine groups **6.D** was synthesised.[22] The self-assembly of **6.D** under the same conditions gave rise to a

product having the formula $Fe^{II}_8L^D_{12}$ by ESI-MS, where L^D represent the ligand incorporating subcomponent **6.D**. The ^{1}H NMR spectrum revealed the presence of a predominant species with three magnetically distinct ligand environments, consistent with tetragonal prismatic structure **6.4** with *mer* coordination at all vertices (Figure 6.4).

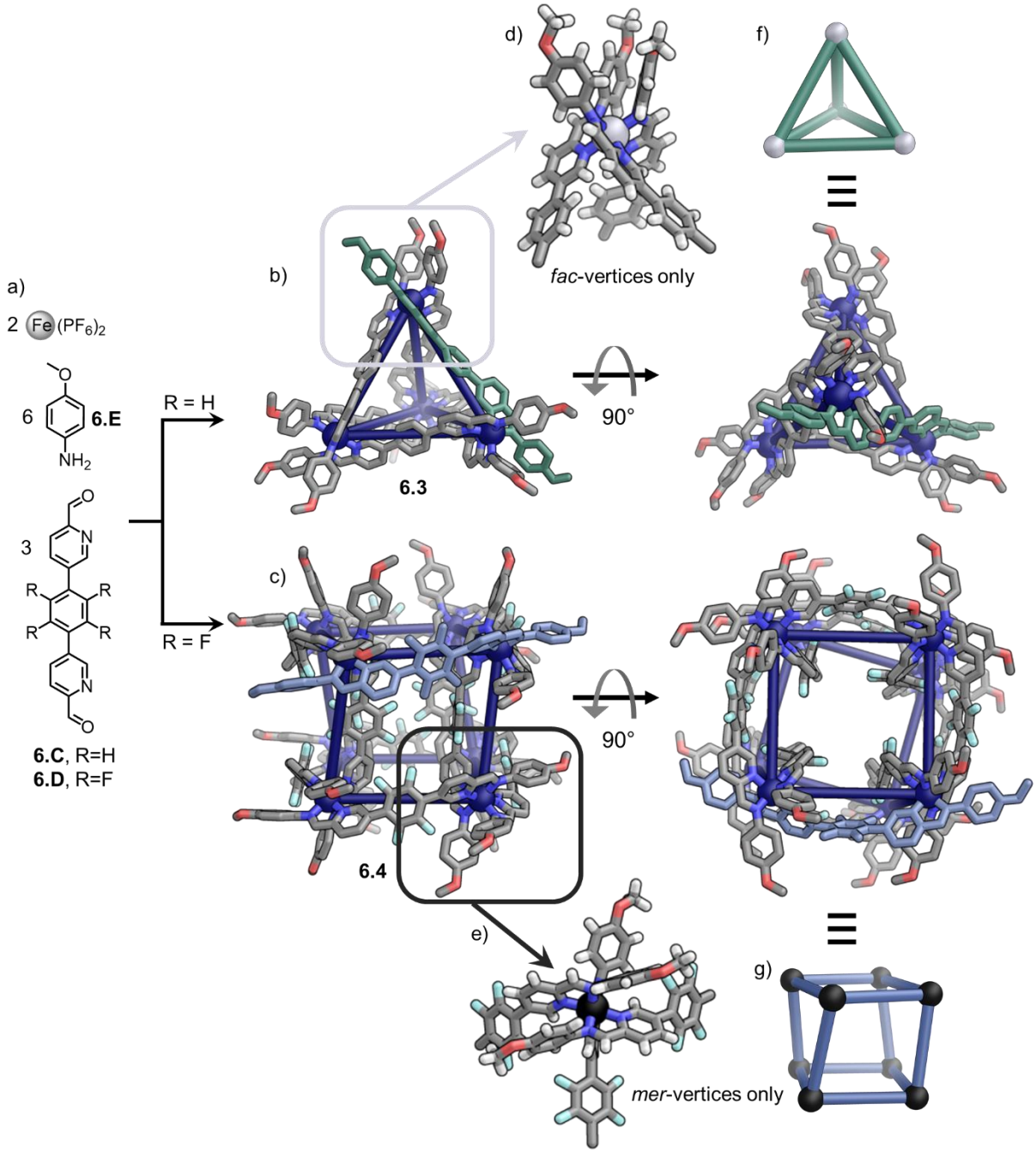


Figure 6.4 | a) Self-assembly of **6.3** and **6.4** from hydrogenated subcomponent **6.C** and fluorinated subcomponent **6.D**, respectively. b and c) Views of the single crystal X-ray structures of **6.3** and **6.4** showing the side view and the top views down the main symmetry axis (C_3 and C_4 respectively), with linkages between Fe^{II} centres (blue spheres) added in order to highlight the framework. d and e) Zoom on the vertices of **6.3** and **6.4**, highlighting the differing nature of the vertices geometry, respectively *fac* or *mer* only. f and g) Schematic representation of **6.3**, **6.4**.

6.1.2 Quantification by mass spectrometry

As subcomponents **6.C** and **6.D** formed contrasting structures with exclusively *fac* (tetrahedron) or *mer* (tetragonal prism) vertices, more information on the factors leading to one or the other structure could be obtained by studying the assemblies formed by mixing the two subcomponents. Due to the similarity in length and shape of the two ligands, a library of heteroleptic assemblies should be formed if neither ligand had a distinct preference for either structural type (Figure 6.1 e).

Preformed tetrahedron **6.3** and tetragonal prism **6.4** were combined in a 2:1 ratio to allow for equimolar amounts of L^C and L^D and were left to equilibrate at 50 °C for three days. The ^{1}H NMR spectrum of the mixture became significantly more complex during this time. Similarly, subcomponents **6.C** and **6.D** were combined in equimolar amount and reacted with **6.E** and $Fe(PF_6)_2$. After equilibration for three days, the same spectra (^{1}H NMR and ESI-MS) as in the first case were obtained. As the non-fluorinated subcomponent **6.C** and fluorinated subcomponent **6.D** were present in a 1:1 ratio and no further changes were observed following longer equilibration times (up to two weeks), we inferred that the system had reached thermodynamic equilibrium. Furthermore, disassembly of the cages did not appear to have a significant energy barrier as both outcomes were identical.

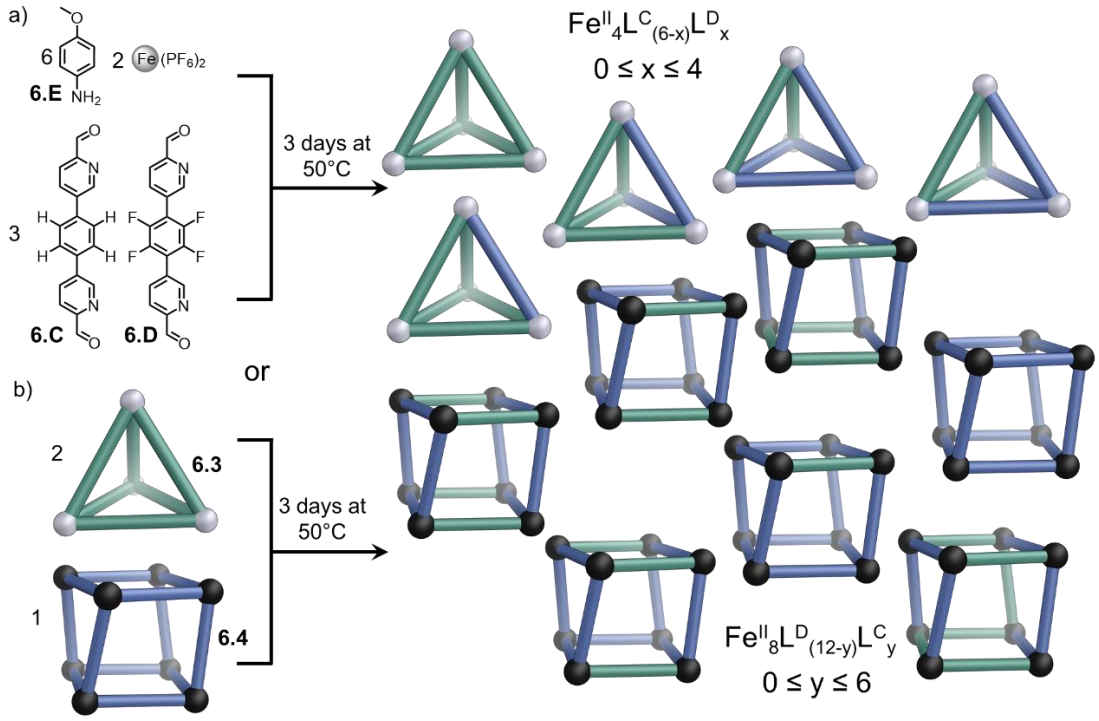


Figure 6.5 | Mixing experiment starting from either subcomponents **6.C** and **6.D** (a) or from the preformed cages **6.3** and **6.4** (b), giving the same library of tetrahedral ($Fe^{II}_4L^C_{(6-x)}L^D_x$) and prismatic ($Fe^{II}_8L^D_{(12-y)}L^C_y$) cages after equilibration at 50°C for 3 days.

The intractability of the ^{1}H NMR spectrum prevented any further analysis via this technique. The broadness was inferred to come from the complexity of the library of cages formed, arising both from the number of species (up to five tetrahedra and 13 prisms) and the different structural arrangements these might adopt.

ESI-MS was therefore used to study the mixture, where each species formed could be observed due to their differing *m/z* ratio. The ESI-MS spectrum indicated the presence of $Fe^{II}_4L^{C}_{(6-x)}L^{D}_x$ tetrahedra incorporating between zero and four fluorinated ligands **6.D**, but no tetrahedra incorporating five or six. Peaks corresponding to $Fe^{II}_8L^{D}_{(12-y)}L^{C}_y$ tetragonal prisms incorporating between zero and six non-fluorinated ligands **6.C** were also observed, but none incorporating more than six (Figure 6.6).

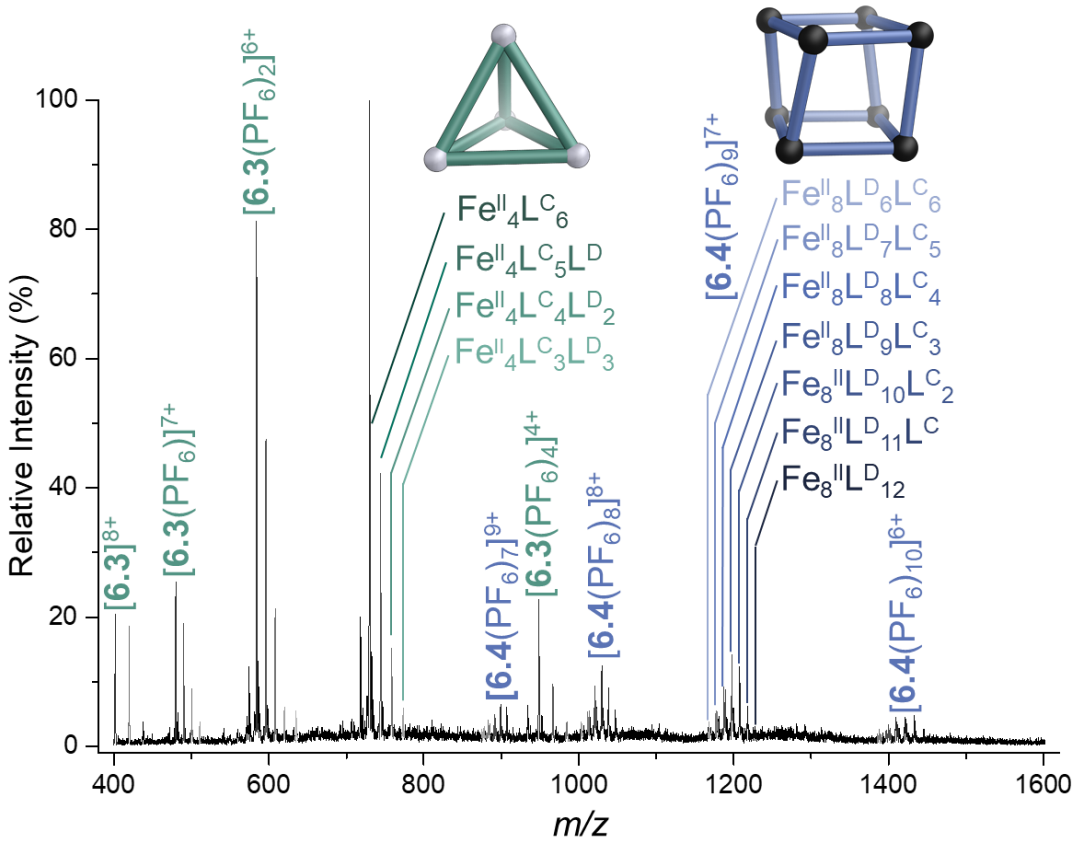


Figure 6.6 | LR-ESI-MS of the mixture after equilibration showing the peaks corresponding to the heteroleptic species $Fe^{II}_4L^{C}_{(6-x)}L^{D}_x$ and $Fe^{II}_8L^{D}_{(12-y)}L^{C}_y$ formed.

When observing a peak corresponding to a single charge state of either architecture, the response factor within a charge state was hypothesised to be independent of the number of fluorinated ligands present. The response factors (related to the peak integrals) were therefore assumed to be independent of the number of fluorinated ligands present in each tetrahedron or prism, or which structural isomer was present. Consequently, the relative concentrations of each

of the congeners was assumed to be proportional to the integral of their *m*/*z* peaks within clusters of peaks corresponding to structures of the same type incorporating different numbers of ligands **6.C** and **6.D**. This method had previously provided consistent results in the context of other complex metallosupramolecular architectures.[35] However, no assumptions on the proportions of one structure type (tetrahedra or prism) compared to the other structure were made based on the likeliness of varying ionization between the two structure types.

Within each charge state, the integrals of the *m*/*z* peaks were normalised so that the sum of the integrals was equal to one. These integrals were averaged across all observed charge states to give the relative amount of each congener in solution. The corresponding statistical distribution was obtained by normalising the binomial distribution (Pascal's triangle with n=6 for the tetrahedron and n=12 for the tetragonal prism) so that the sum of the binomial coefficients was equal to one. The observed values were plotted alongside the binomial distribution (Figure 6.7). Any deviation from the entropically-favoured statistical (binomial) distribution of heteroleptic assemblies, even in the absence of complete narcissistic self-sorting, would indicate a preference for the ligands to self-sort into their initial structural type.[36-39] Here, a clear deviation from the binomial distribution for both tetrahedra and tetragonal prisms was observed across both architectures (Figure 6.7). These deviations reflected the energetic preference of non-fluorinated subcomponent **6.C** to form tetrahedra and fluorinated subcomponent **6.D** to form tetragonal prisms.

The energetic preferences could be further quantified by establishing a set of equilibrium constants between congeners. Within each series (tetrahedral or prismatic) the ratio between the observed proportion of each species and its expected proportion, based upon a binomial distribution was calculated. Each congener within a series was identified by the number of fluorinated ligands *N* it possesses. The relative Gibbs energy of each congener was then determined from these equilibrium constants using Equation 6.1.

Equation 6.1 $$\boldsymbol{\Delta E_{rN} = -RT\ln(K_N)}, \quad \textbf{with } \boldsymbol{K_N = \frac{\left(\frac{I_N}{I_{statN}}\right)}{I_0}}$$

where I_N represents the normalised integral of the mass peak for the structure incorporating *N* fluorinated ligands, I_{statN} the corresponding normalised binomial coefficient (expected normalised integral for the structure incorporating *N* fluorinated ligands if there was no bias), and I_0 the normalised integral of the mass peak for the original structure (**6.3** or **6.4**).

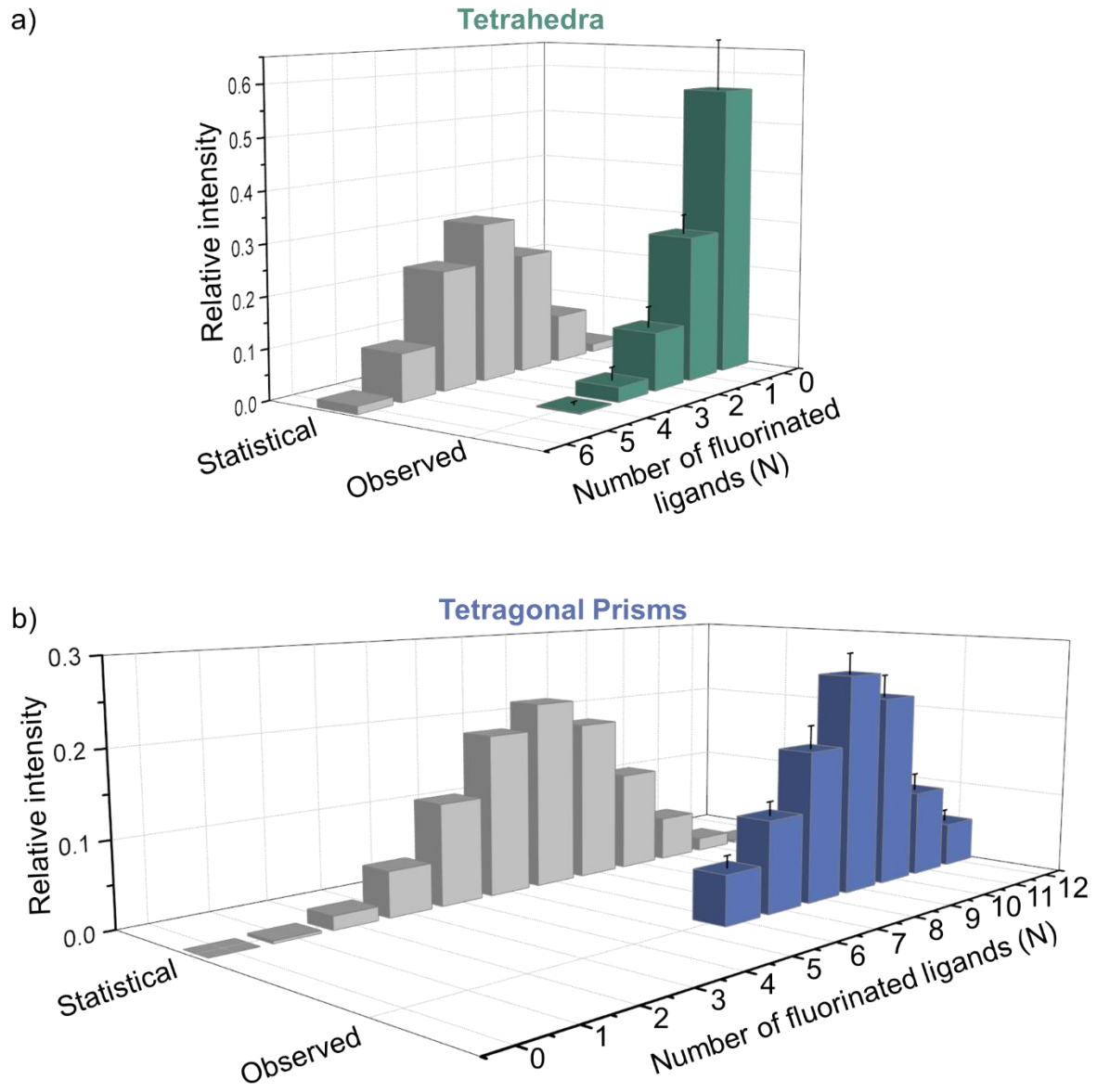


Figure 6.7 | Proportion of each species observed for (a) the family of tetrahedra derived from **6.3** (green) and (b) the family of tetragonal prisms derived from **6.4** (blue) compared to the binomial (statistical) distribution (grey). Error bars represent the standard deviations of the amounts of each congener measured between the different charge states observed in the ESI-MS.

The Gibbs energies followed a linear trend which was fitted to a linear least-squares model. For each fluorinated subcomponent **6.D** incorporated into the tetrahedral cage, the structure was destabilised by an average of 4.1 kJ mol^{-1} (Figure 6.8 a). For each non-fluorinated subcomponent **6.C** incorporated into the tetragonal prism, the structure was destabilised by an average of 2.7 kJ mol^{-1} (Figure 6.8 b). These are weighted average energy for all structural isomers, as it is impossible to differentiate between structural isomers *via* ESI-MS.

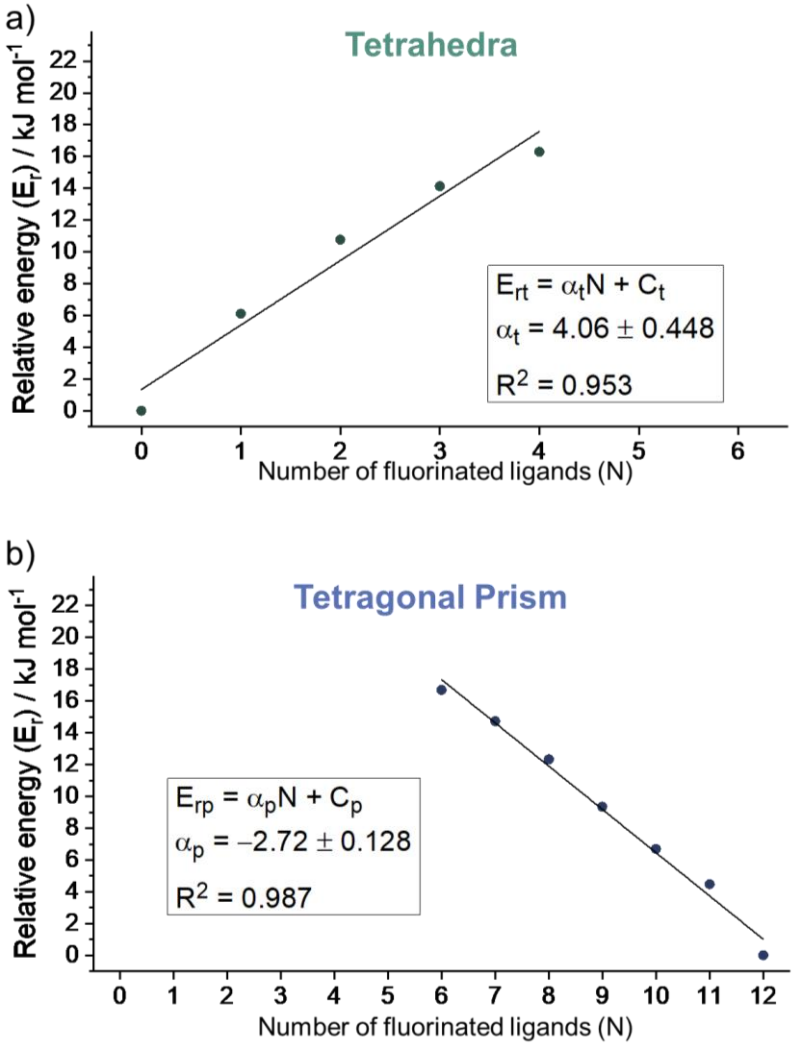


Figure 6.8 | Plots of the energy of each species relative to a) tetrahedron **6.3**, containing only non-fluorinated ligand **6.C**, and b) the tetragonal prism **6.4**, containing only fluorinated ligand **6.D**. E_{rt} and E_{rp} refer to the relative energies between congeners respectively in the cases of the tetrahedral series and the prismatic series. α represents the energetic destabilization per ligand exchanged.

From these values, it was established that it is more energetically costly to incorporate a fluorinated ligand into a tetrahedron than a non-fluorinated ligand into a tetragonal prism. This is consistent with the observation made on the mononuclear complexes. Complex **6.1** formed as a statistical *fac*-*mer* mixture, whereas *mer*-**6.2** formed stereoselectively. Entropy tends to favour the formation of structures with the minimum number of components.[40-41] The tetrahedron incorporates half the number of building blocks compared to the prism in this study. As the non-fluorinated mononuclear complex **6.1** has no preference for *fac* or *mer* geometry, entropy drives the system towards the formation of the smaller tetrahedron. The fluorinated ligands form exclusively vertices of *mer*-geometry, leading to the formation of larger prismatic structures. This suggests that prisms are enthalpically favoured, outweighing the entropic preference to form tetrahedra.

6.3 Sorting regimes and composition of heteroleptic mixtures

Given the many uses of self-assembled cages, the ability to gauge the composition of heteroleptic mixtures of capsules could be of high value. Understanding the factors (electronics, flexibility, size and shape) influencing the incorporation of different ligands into heteroleptic species could allow better control over the outcome of the self-assembly process.

6.3.1 Chemical system

Previously, Rizzuto *et al.* demonstrated that subcomponent self-assembly of triamines such as **6.H** and **6.I** with 2-formylphenanthroline **6.G** and either Cd^{II} or Zn^{II} formed a series of supramolecular $M^{II}_6L_4$ *pseudo*-octahedra.[42] However, the formation of discrete species was not observed when larger triamines such as **6.J** and **6.K** were employed. Co^{II} has proven useful in stabilizing larger and more flexible supramolecular architectures in the past.[43-46] This phenomenon can be explained by the slower exchange kinetics and flexible coordination sphere of this metal ion. It was thus hypothesised that employing Co^{II} with larger triamines and **6.G** could stabilise the formation of larger octahedra.

The self-assembly of **6.G** (12 eq.), $Co^{II}(OTf)_2$ (6 eq.) and triamines **6.H – 6.K** (4 eq.) carried out by Felix Rizzuto resulted in the formation of assemblies **6.5 – 6.8** after heating at 70 °C for 16 h (Figure 6.9). ESI-MS confirmed the $Co^{II}_6L_4$ stoichiometry in all cases. The ^{1}H NMR spectra of **6.5 – 6.7** were highly symmetric, consistent with the threefold symmetry of the ligands and the overall *T* symmetries of the expected octahedra.

The ^{1}H NMR spectrum of **6.8**, was more complex with three sets of ligand signals instead of the single one expected for the *T* symmetric cage. Single crystal X-ray diffraction studies carried out by Felix Rizzuto revealed a structure where each of the faces of the polyhedron corresponded to a scalene triangle. The alternating metal corners around the equatorial belt are forced out of the plane by the ligands, leading to their scalenohedral arrangement. Edge-to-face aromatic interactions between the triazine rings of the ligand and adjacent phenanthroline moieties was inferred to reinforce the configuration, as seen in the X-ray structure. Furthermore, the high electron density of **6.K** was hypothesised to contribute to the edge-to-face aromatic interactions in **6.8**. This would explain the differences in the structures formed by **6.K** as opposed to **6.J**, which possesses similar geometry and size but different electronic properties.

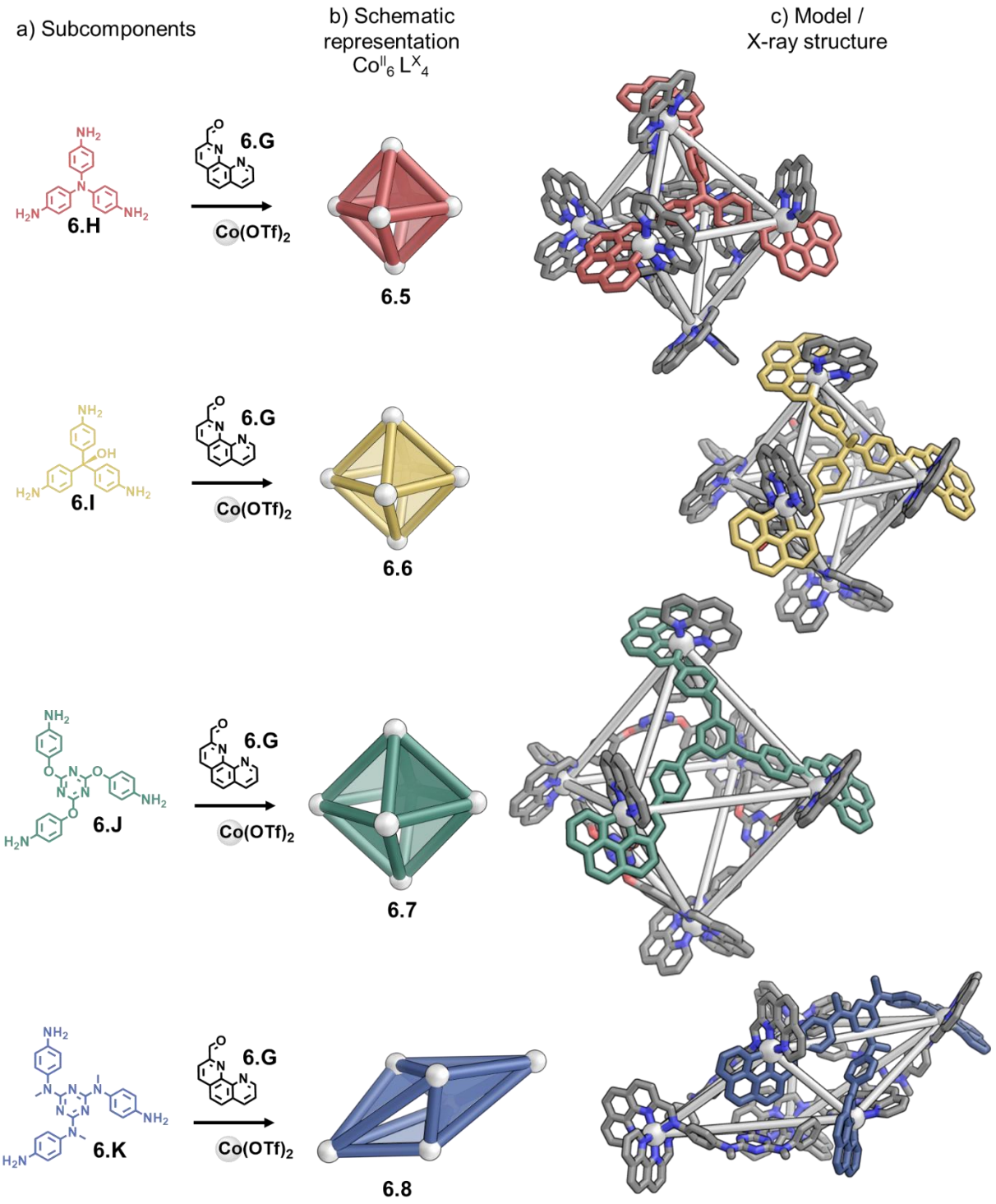


Figure 6.9 | Formation of supramolecular cages **6.5 – 6.8** (b – schematic representation of the structures) from a) subcomponent **6.G**, $Co(OTf)_2$ and tiamines **6.H** – **6.K**, respectively. c) Single crystal X-ray structures of **6.5,** and **6.8** and MM3-optimised models of **6.6** (based on the X-ray structure of the analogue with Cd) and **6.7**. Linkages between Co^{II} centres (white spheres) added to highlight the framework, hydrogens are omitted for clarity.

6.3.2 Quantification by mass spectrometry

To probe the effect of the ligands' properties (size, shape, flexibility, electronic properties) on self-sorting regimes, triamines **6.H – 6.K** were self-assembled two by two with **6.G** and $Co(OTf_2)_2$ and the outcome of the mixing experiment was assessed by NMR and MS.

Two triamines (1 eq. of each), 2-formylphenanthroline **6.G** (6 eq.) and $Co(OTf)_2$ (3 eq.) were combined in CH_3CN and heated at 70 °C for 18 h. Stock solutions of homoleptic species **6.5**, **6.6**, **6.7** and **6.8** were also prepared. The same amount of two homoleptic cage solutions were combined and heated at 70 °C for 18 h. ^{1}H NMR and ESI-MS were collected after cooling to room temperature. Both methods gave similar spectra that did not change further upon longer equilibration times (up to one week). Hence, the system was assumed to be at thermodynamic equilibrium in both cases. For ease of preparation and to limit errors due to weighing small amounts of material, the method involving stock solutions was used throughout the rest of this work.

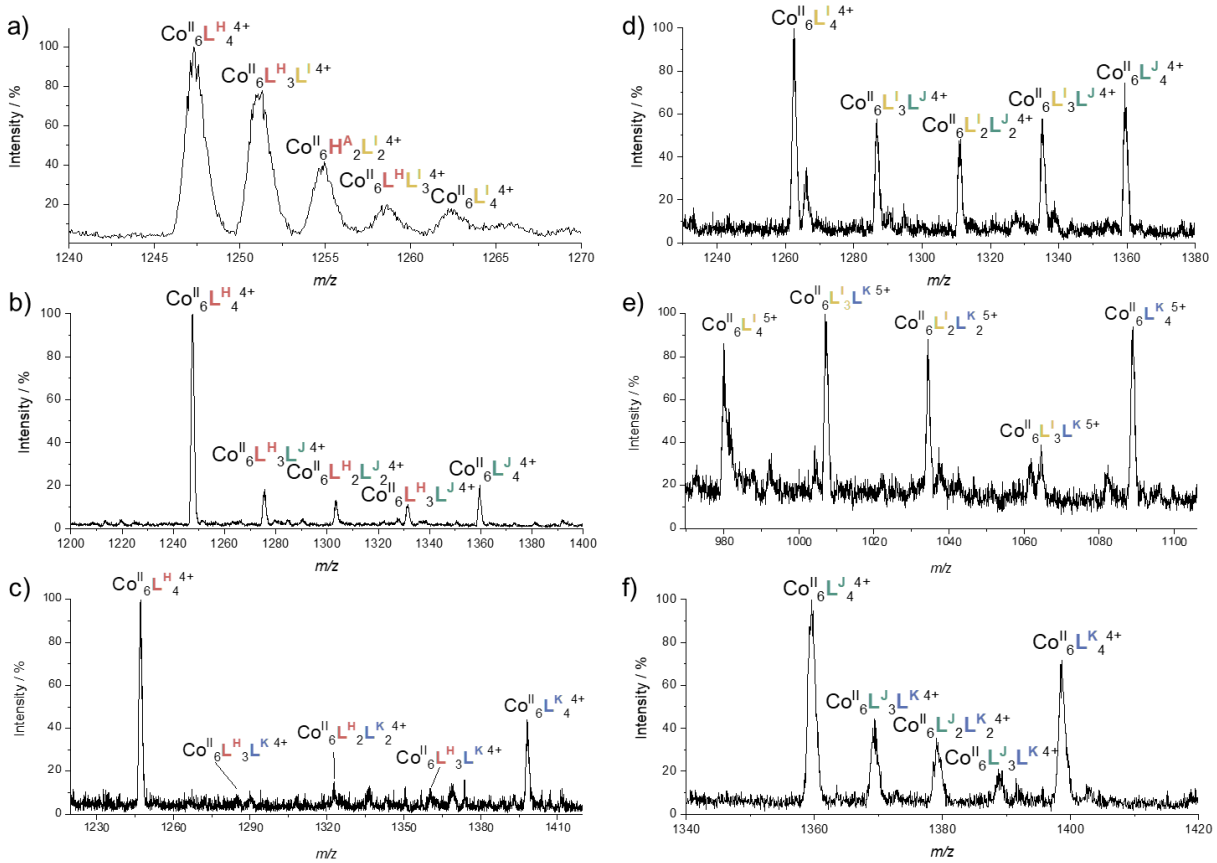


Figure 6.10 | LR-ESI-MS of library of cages of the formulae $Co^{II}_6L^Y_xL^Z_{4-x}$ obtained by mixing a) **6.5** and **6.6**, b) **6.5** and **6.7**, c) **6.5** and **6.8**, d) **6.6** and **6.7**, e) **6.6** and **6.8** and f) **6.7** and **6.8**.

ESI-MS was used to probe the effect of each ligand on the outcome of the self-sorting process. In most cases, the mixtures containing two different subcomponents were observed to socially self-sort into $Co^{II}_6L^Y_xL^Z_{4-x}$ architectures by ESI-MS (where L^Y and L^Z represent ligands incorporating subcomponents **6.Y** and **6.Z**, respectively, Figure 6.10). Only the mixture of **6.H** and **6.K** was observed to undergo near-narcissistic sorting where the main species observed were the homoleptic **6.5** and **6.8**. All other combinations of tiamines (**6.H** and **6.I**, **6.H** and **6.J**, **6.I** and **6.J** as well as **6.J** and **6.K**) led to the formation of all possible products. Ligands formed from **6.J** and **6.K** are significantly larger than **6.H** and **6.I**, indicating that ligand geometries and size did not have to match closely to form heteroleptic structures. MM3 molecular models of the $Co^{II}_6L^I_xL^J_{4-x}$ complexes calculated by Felix Rizzuto indicated that these heteroleptic structures did not suffer from significant distortions of the ligands or the coordination environments of the Co^{II} ions.[23]

The ionisation and level of detection in the mass spectrometer is very likely to differ greatly for each species. However, information on the product distribution between homoleptic and heteroleptic could be obtained for each experiment. To do so, freshly prepared mixtures of preformed cages (non-equilibrated) at equal concentration were analysed in parallel to the equilibrated ones and compared to each other. For example, a fresh mixture of **6.5** and **6.6** was compared against equilibrated $Co^{II}_6L^H_xL^I_{4-x}$. Peaks corresponding to both homoleptic and heteroleptic species were observed in the mass spectra of equilibrated mixtures, whereas no peaks matching with heteroleptic species were observed in the freshly combined mixtures. This allowed the relative response factor of the heteroleptic species by themselves to be gauged.

For all mixtures, the +4, +5 and +6 charge states were used for analysis as lower charge states were outside the spectral window of the instrument, and higher charge states overlapped with too many fragments to be incorporated into the analysis. Each region of the mass spectrum (corresponding to either a homoleptic or heteroleptic cage) was integrated for both equilibrated and fresh mixtures. To compensate for overlapping peaks for fragments in the heteroleptic regions, the integrals for regions corresponding to heteroleptic cages in freshly combined mixtures were subtracted from the integrals of the corresponding regions in the spectra of the equilibrated mixtures. No free ligand was observed in the NMR spectrum, hence the assumption that all homoleptic species consumed are converted into heteroleptic species was made. Under this assumption, integrals for each charge state could be normalised to unity to facilitate comparison (Figure 6.12).

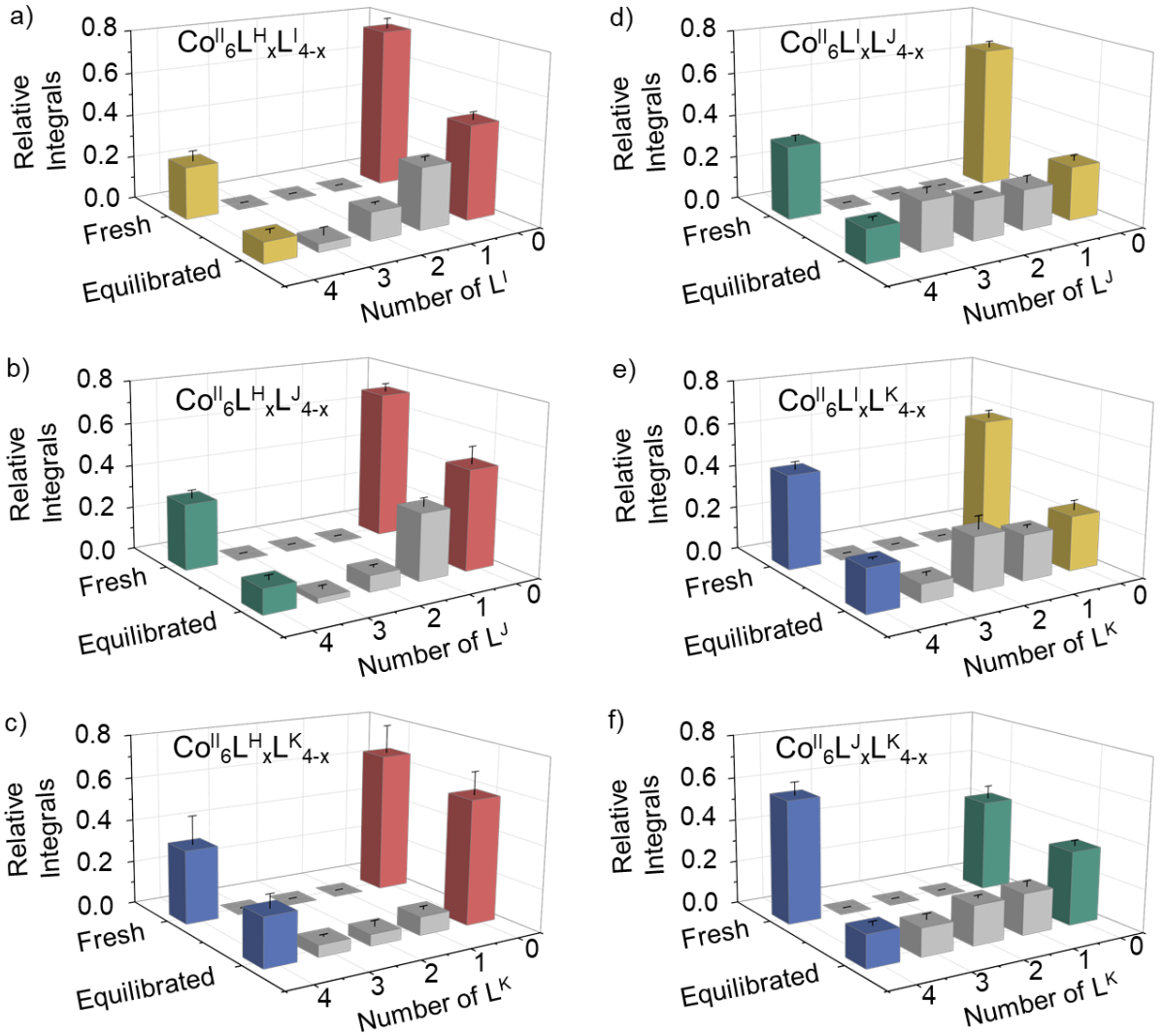


Figure 6.11 | Plot of the normalised integrals for either fresh or equilibrated mixtures of a) **6.5** (red) and **6.6** (yellow), b) **6.5** (red) and **6.7** (green), c) **6.5** (red) and **6.8** (blue), d) **6.6** (yellow) and **6.7** (green), e) **6.6** (yellow) and **6.8** (blue), and f) **6.7** (green) and **6.8** (blue). The heteroleptic species $Co^{II}_6L^Y_xL^Z_{4-x}$ (x = 1, 2 or 3) formed are shown in grey. Error bars represent the standard deviations for each species observed in the ESI-mass spectra across the different charge states.

No direct comparison between structures incorporating different ratios of ligands could be made in this case as the response factors of the homoleptic species were too different, and thus it was inferred that the response factors of heteroleptic species will vary as well. However, comparison between the overall amounts of homoleptic species in both equilibrated and fresh mixtures can be made, as they are structurally identical, and thus exhibit identical ionisation characteristics. This method eliminated effects associated with different detection intensities of charged complexes with different structures and different ligand configurations.

The relative decrease in the integral of peaks attributed to homoleptic species was used to quantify the amount of each ligand integrated into all heteroleptic species. For example, in a

self-sorted mixture of **6.H** and **6.I**, the percentage of **6.5** and **6.6** remaining in the equilibrated mixture, as compared to **6.5** and **6.6** in the fresh mixture, was directly linked to the proportion of homoleptic species consumed by the generation of heteroleptic species. The decrease in the overall amount of homoleptic species thus relates to the percentage of heteroleptic species formed in solution. Likewise, the individual percentage decrease in either **6.5** or **6.6** indicates the proportion of subcomponent **6.H** or **6.I** respectively integrated into heteroleptic cages.

Hence, the percentage of homoleptic species remaining after equilibration of the mixtures is obtained by comparing the normalised integrals of identical homoleptic species in the equilibrated ($I_{x(eq)}$) and the fresh mixtures ($I_{x(fresh)}$). The percentage of each homoleptic species (**X** = **6.5**, **6.6, 6.7**, or **6.8**) remaining in the equilibrated mixtures can be calculated using Equation 6.2.

Equation 6.2

$$\%\mathbf{X(eq)} = \frac{I_{\mathbf{X(eq)}}}{I_{\mathbf{X(fresh)}}} * 100$$

The values obtained for all mixtures are given in Table 6.1.

Table 6.1 | Percentage of homoleptic species remaining after equilibration of the mixtures of homoleptic cages.

Mixture	*Percentage of homoleptic species remaining*	
6.5 and **6.6**	Percentage of **6.5**	Percentage of **6.6**
	59 ± 3	40 ± 8
6.5 and **6.7**	Percentage of **6.5**	Percentage of **6.7**
	68 ± 8	37 ± 2
6.5 and **6.8**	Percentage of **6.5**	Percentage of **6.8**
	89 ± 7	69 ± 11
6.6 and **6.7**	Percentage of **6.6**	Percentage of **6.7**
	38 ± 5	46 ± 9
6.6 and **6.8**	Percentage of **6.6**	Percentage of **6.8**
	45 ± 5	46 ± 5
6.7 and **6.8**	Percentage of **6.7**	Percentage of **6.8**
	81 ± 11	27 ± 4

The distribution between heteroleptic and homoleptic species is obtained by measuring the overall percentage of homoleptic species remaining after equilibration (Equation 6.3). The quantity of heteroleptic species is the reciprocal of this percentage (Equation 6.4). The distribution between species in a mixture of cages **X** and **Y** is given by:

Equation 6.3 $$\%Homoleptic(\boldsymbol{X},\boldsymbol{Y}) = \frac{\%\boldsymbol{X}(\mathbf{eq})+\%\boldsymbol{Y}(\mathbf{eq})}{2}$$

Equation 6.4 $$\%Heteroleptic(\boldsymbol{X},\boldsymbol{Y}) = 100 - \%Homoleptic$$

The values obtained for all mixtures are given in Table 6.2.

Table 6.2 | Distribution between homoleptic and heteroleptic species after equilibration of the mixtures of homoleptic cages.

Mixture	*Percentage of homoleptic species*	*Percentage of heteroleptic species*
6.5 and **6.6**	49 ± 6	51 ± 6
6.5 and **6.7**	53 ± 5	47 ± 5
6.5 and **6.8**	79 ± 9	21 ± 9
6.6 and **6.7**	42 ± 7	58 ± 7
6.6 and **6.8**	46 ± 5	54 ± 5
6.7 and **6.8**	54 ± 7	46 ± 7

Under the assumption that all homoleptic species consumed are converted into heteroleptic species during equilibration, the percentage of ligand **6.H**, **6.I**, **6.J** or **6.K** incorporated into the heteroleptic cages of the equilibrated mixtures can thus be quantified. This quantity corresponds to the percentage decrease of the corresponding homoleptic species in the equilibrated mixture, as compared to that in the fresh mixture. In a mixture of cages **X** and **Y**, the percentage of L^x incorporated in the heteroleptic species can be calculated using Equation 6.5.

Equation 6.5 $$\%L^x(hetero) = \frac{(100-\%\mathbf{X}(\mathbf{eq}))}{(100-\%\mathbf{X}(\mathbf{eq}))+(100-\%\mathbf{Y}(\mathbf{eq}))} * \%Heteroleptic(\boldsymbol{X},\boldsymbol{Y})$$

The ligand composition of heteroleptic cages can be obtained from the percentage of each ligand incorporated into heteroleptic species by cross multiplication. In a mixture of **X** and **Y**, the ligand composition L^x of heteroleptic cages can be calculated using Equation 6.6.

Equation 6.6 $$\%L^{x}(composition) = \frac{\%\mathbf{X}(hetero)}{\%\mathbf{X}(hetero)+\%\mathbf{Y}(hetero)} * 100$$

The values obtained for all mixtures are given in Table 6.3.

Table 6.3 | Ligand composition of heteroleptic species after equilibration of the mixtures of homoleptic cages.

Mixture	*Ligand composition of heteroleptic species*	
6.5 and **6.6**	Percentage of **6.H**	Percentage of **6.I**
	41 ± 3	59 ± 8
6.5 and **6.7**	Percentage of **6.H**	Percentage of **6.J**
	34 ± 8	66 ± 2
6.5 and **6.8**	Percentage of **6.H**	Percentage of **6.K**
	27 ± 17	73 ± 25
6.6 and **6.7**	Percentage of **6.I**	Percentage of **6.J**
	53 ± 4	47 ± 7
6.6 and **6.8**	Percentage of **6.I**	Percentage of **6.K**
	50 ± 5	50 ± 5
6.7 and **6.8**	Percentage of **6.J**	Percentage of **6.K**
	21 ± 12	79 ± 4

This method thus provided information on both the product distribution within the sorted mixtures (*i.e.*, the percentage decrease observed for homoleptic species), as well as the proportion of each ligand integrated into the heteroleptic species for each experiment. The results are summarised in Figure 6.12.

If there was no energetic bias for the formation of homoleptic *vs* heteroleptic structures, a socially-sorted mixture would be obtained, consisting of 12.5% of homoleptic species vs. 87.5% of heteroleptic species, following a binomial distribution. All sorting experiments were observed to deviate substantially from these values (Figure 6.12, left). When heteroleptic species were formed (all cases but the combination of **6.5** and **6.8** which followed near-narcissistic sorting), the makeup was approximately 1:1 homoleptic:heteroleptic. This indicated a preference for the formation of homoleptic species, which can be attributed to the strain incorporated into mixed-ligand species. Indeed, all ligands have different sizes, leading to mismatches when forming heteroleptic structures. Similarly, in the absence of enthalpic bias, the ligand composition of the heteroleptic species should be even (1:1). However, in some

cases, higher proportions of one ligand over another were integrated into heteroleptic structures (Figure 6.12, right). This indicated that some cage frameworks preferentially incorporated specific ligands into their structures.

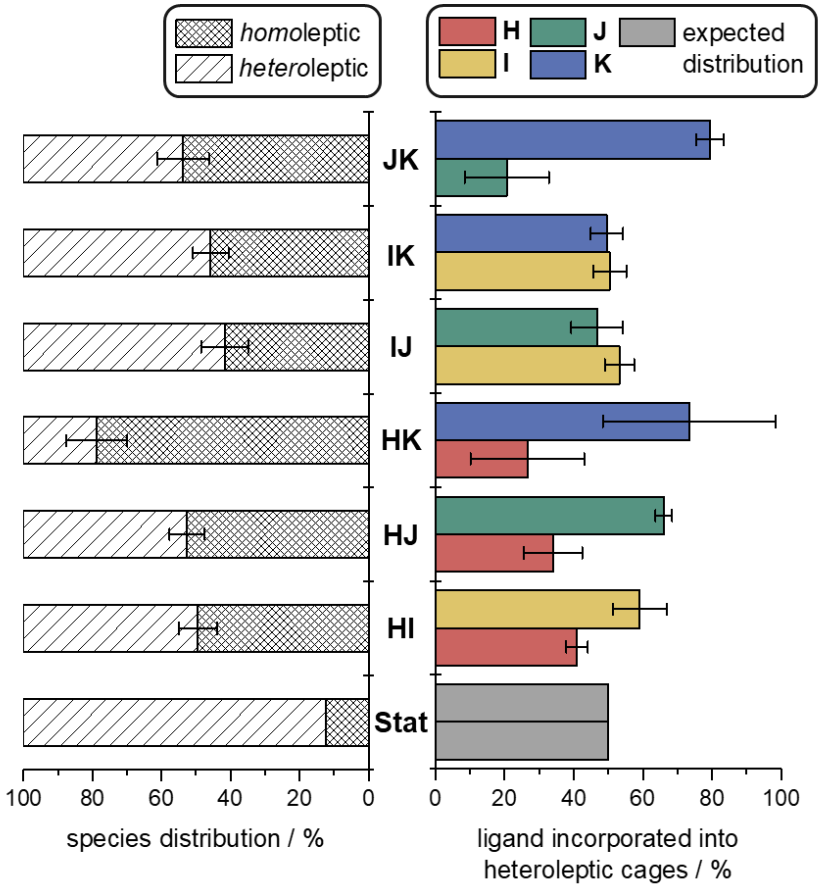


Figure 6.12 | Product distributions determined from ESI-MS experiments: percentage distribution of homo- vs heteroleptic products (left) and the ligand make-up of the heteroleptic species (right), compared to the statistical binomial distribution (Stat).

In the case of the mixtures $Co^{II}_6L^H_xL^Y_{4-x}$ formed from **6.H** and any of the other subcomponents, only small amounts of **6.H** were incorporated into heteroleptic species, and homoleptic species were favoured over heteroleptic. On the other hand, when **6.I** was employed with **6.K** or **6.J**, equal amounts of homoleptic and heteroleptic species were formed, which incorporated both ligands in similar proportions. This demonstrates that the slight difference of flexibility between planar **6.H**, and pyramidal **6.I** has a considerable impact on the outcome of the sorting. Furthermore, the flexibility of the ether linkages in **6.J** and amine linkages in **6.K** appears to enable geometrical diversity, allowing different conformations to be adopted in order to accommodate the differently-sized subcomponents. However, in the case of $Co^{II}_6L^J_xL^K_{4-x}$, where the triamines were the same size, a clear bias towards the integration of **6.K** over **6.J** into heteroleptic species was observed.

To further explain these observations, the heteroleptic mixtures were studied by IM-MS (ion-mobility mass spectrometry). Although all structures are of identical stoichiometry, they are not isostructural and the difference in size between octahedron and scalenohedron should result in different drift times detected by the IM-MS instrument. Indeed, when species with identical stoichiometry are ionised, they can be differentiated by the time it takes them to travel through a buffer gas under an electric field.[47] Larger structures have a greater number of interactions with the buffer gas in the ion-mobility chamber leading to extended or elongated structures drifting slower than smaller structures of the same charge.[48-49] The drift time of each supramolecular species is thus directly related to its collisional cross section, which in turn depends on its geometry. Supramolecular architectures have previously been successfully studied with this technique.[50] IM-MS data was acquired and analysed for the mixtures of **6.H** and **6.K** (Figure 6.13), **6.I** and **6.K** (Figure 6.14) and **6.J** and **6.K** (Figure 6.15).

The analysis of $Co^{II}_6L^{H}_xL^{K}_{4-x}$ revealed that the *m/z* regions corresponding to the purely homoleptic cages **6.5** and **6.8** had unique drift times (9.6 – 10.4 ms and 10.4 – 11.0 ms, respectively) which were each attributed to the octahedral and scalenohedral structures (Figure 6.13 c and d). Unfortunately, peaks for the heteroleptic species were too weak to examine, due to the near-narcissistic sorting of this system.

When $Co^{II}_6L^{I}_xL^{K}_{4-x}$ was analysed by IM-MS, two distinct drift times were observed at 9.6 – 10.6 ms and 10.6 – 11.2 ms, corresponding to octahedral and scalenohedral geometries respectively. The shorter drift time regime included **6.6** and all heteroleptic species of this mixture, revealing the octahedral geometry of these structures (Figure 6.14 c). Only homoleptic **6.8** was scalenohedral (Figure 6.14 d) showing that **6.I** promoted the formation of heteroleptic octahedra exclusively when combined with **6.K**. This situation was inferred to arise from the impossibility of rigid **6.I** to accommodate the strain necessary to form the scalenohedron, whereas more flexible **6.K** can adopt the right configuration to form an octahedron.

The analysis of $Co^{II}_6L^{J}_xL^{K}_{4-x}$ likewise revealed two distinct, broad drift time regions at 10.0 – 10.5 ms and 10.6 – 11.2 ms attributed to octahedra and scalenohedra respectively. When each region was examined separately, the octahedra region was revealed to be comprised of mainly homoleptic **6.7** and a small proportion of species with higher numbers of **6.J** incorporated (Figure 6.15 c). In contrast, the scalenohedra range was comprised of **6.8**, and most of the heteroleptic species which incorporated a higher proportion of **6.K** (Figure 6.15 d). This data indicated that the scalenohedral geometry is preferred over the octahedral geometry

when the ligands are similarly-sized. This bias was attributed to the favourable edge-to-face aromatic interactions within congeners of **6.8** incorporating higher numbers of the triamine **6.K**.

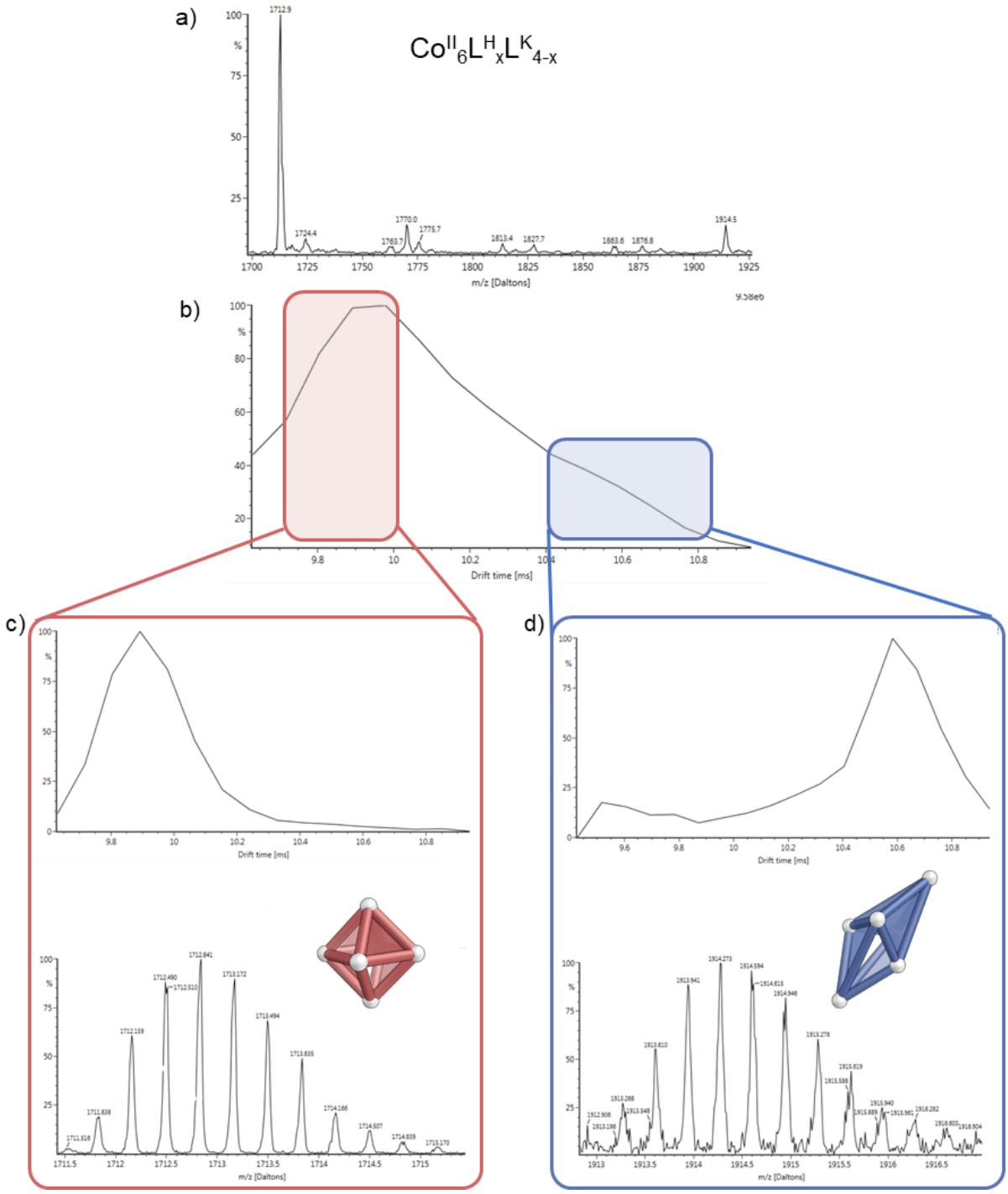


Figure 6.13 | IM-MS data for the mixture generated by combining **6.5** and **6.8** showing a) the *z* = +3 peaks and b) the two broad, overlapping drift time regions at 9.6 – 10.4 ms and 10.4 – 11.0 ms. The *m/z* regions of the purely homoleptic cages **6.5** and **6.8** revealed two unique, separate drift times. Peaks attributed to heteroleptic species were too weak in to examine, due to the near-narcissistic sorting of this system.

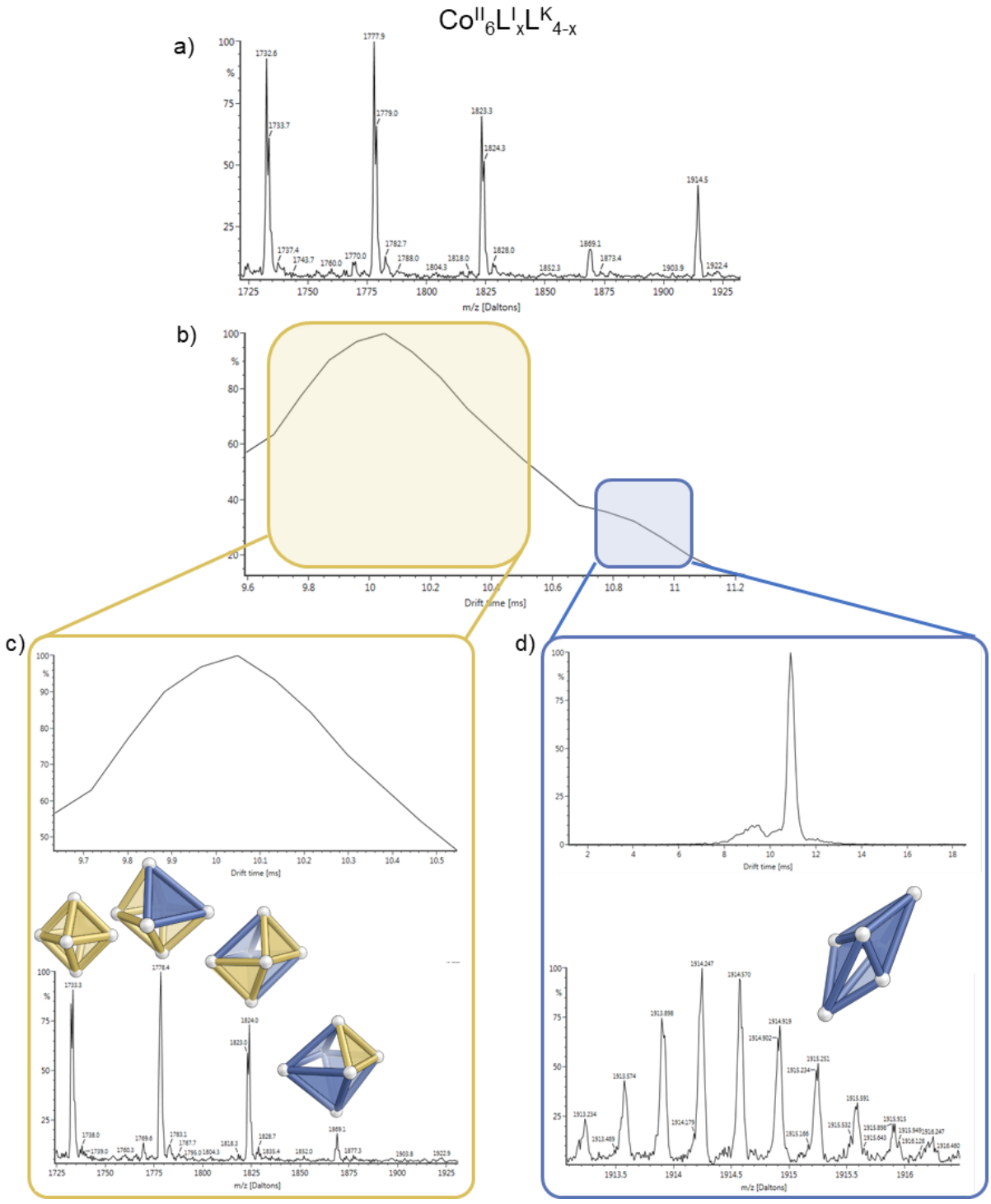


Figure 6.14 | IM-MS data for the mixture generated by combining **6.6** and **6.8** showing a) the *z* = +3 peaks and b) the two distinct, broad drift time regions at 9.6 – 10.6 ms (octahedra) and 10.6 – 11.2 ms (scalenohedra). Separate

examination of the two regions showed that species comprising any number of **6.I** were octahedral (c) and that only **6.8**, comprised entirely of **6.K** was scalenohedral (d).

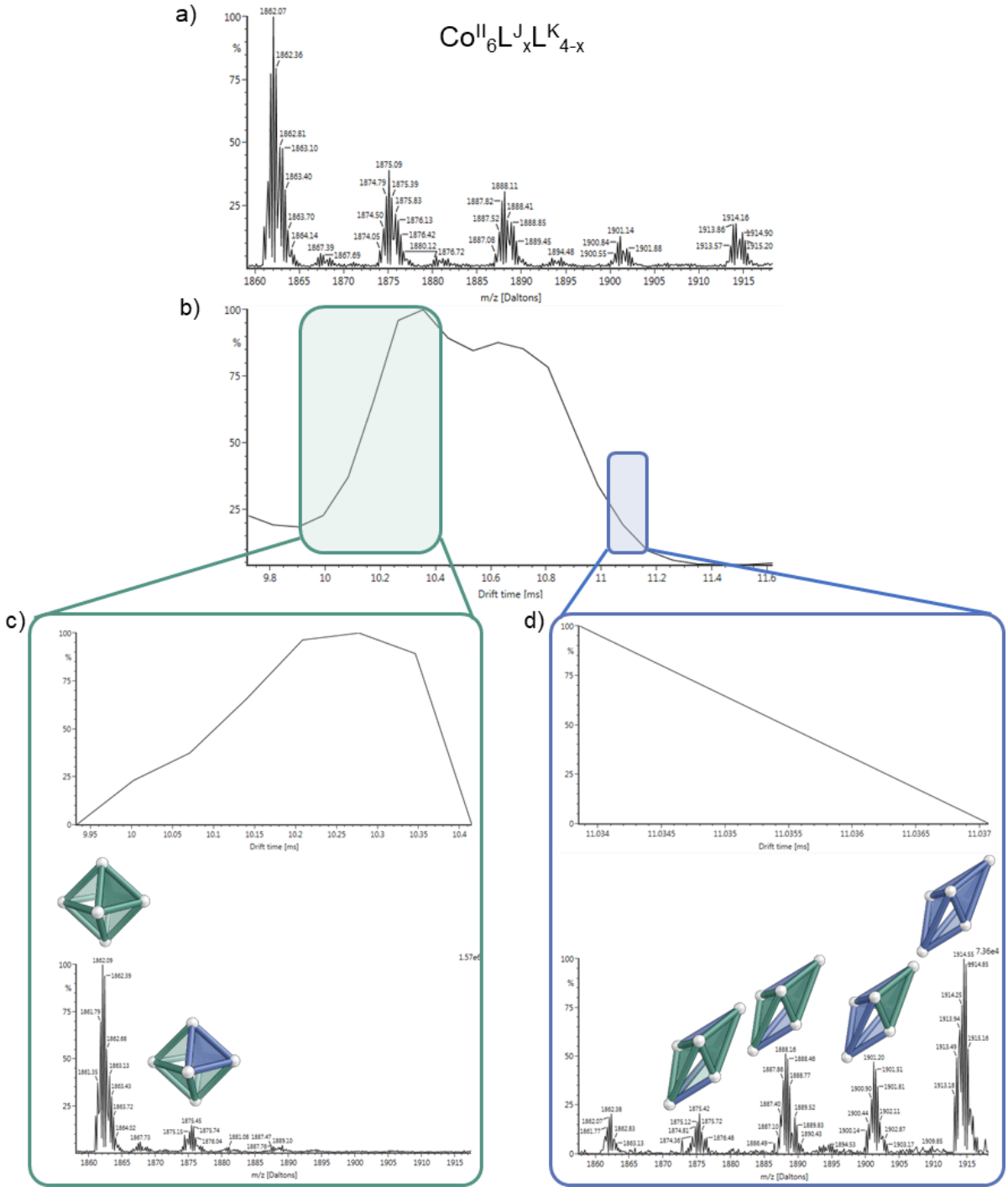


Figure 6.15 | IM-MS data for the mixture generated by combining **6.7** and **6.8** showing a) the *z* = +3 peaks and b) the two distinct, broad drift time regions at 10.0 – 10.5 ms (octahedra) and 10.6 – 11.2 ms (scalenohedra). Separate

examination of the two regions showed that the octahedra region was comprised of species with a higher proportion of **6.J** (c) and that the scalenohedron range was comprised of species with a higher proportion of **6.K** (d).

Combining both ESI-MS and IM-MS techniques, the distinct behaviour of each mixture could be classified and a sorting regime was attributed to each mixture (Figure 6.16). Depending on their size, shape, flexibility, and electronic properties, three distinct sorting regimes were achieved. Subcomponents **6.H** and **6.K** underwent near-narcissistic sorting whereas **6.H** and **6.J**, **6.H** and **6.I** or **6.I** and **6.J** underwent social self-sorting. Finally, geometrically-biased sorting was observed for **6.I** and **6.K** as well as **6.J** and **6.K**, towards octahedra or scalenohedra respectively.

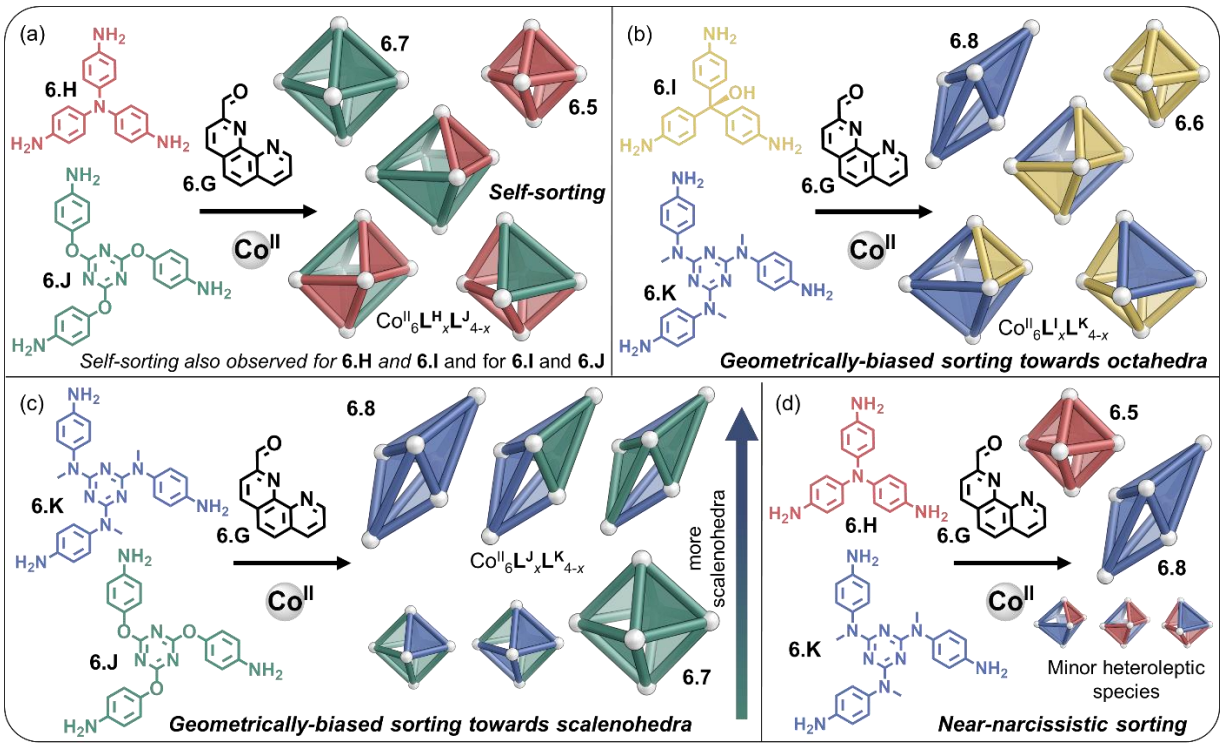


Figure 6.16 | Distinct sorting regimes of the mixing experiments on pairs of amines (**6.H – 6.K**). a) Both homo- and heteroleptic species were observed to form from **6.H** and **6.J**, **6.H** and **6.I** or **6.I** and **6.J**. b) Geometrically-biased sorting where all heteroleptic species formed by **6.I** and **6.K** were octahedral and only the homoleptic **6.8** was scalenohedral. c) Geometrically-biased sorting where structures with a greater proportion of **6.K** formed the scalenohedral shape preferentially for mixtures of **6.J** and **6.K**. d) Near-narcissistic sorting was observed for **6.H** and **6.K**.

6.4 Species amplification by guest binding

The incorporation of molecules able to interact with congeners within a self-sorted library, such as templates, can change the outcome of the sorting process and redistribute the products obtained. Being able to quantify the effect of these external trigger molecules on the library could allow us to gain better knowledge on the synergy that exist within a self-sorted system.

6.4.1 Chemical system

Bilbeisi *et al.* previously reported the formation of $Fe^{II}_4L_4$ tetrahedral cages from **6.H** or **6.I**, picolinaldehyde and $Fe(OTf)_2$ in CH_3CN.[51] No guest encapsulation was observed for these cages, due to the small size of their cavities. The use of Co^{II} had previously proven to increase the cage's cavity when used in self-assemblies instead of smaller coordination sphere metal cations.[52-53] Preliminary results by Rana Bilbeisi showed that the self-assembly of picolinaldehyde (12 eq.) and $Co(NTf_2)_2$ (6 eq.) with either triamine **6.H** or **6.I** (4 eq.) resulted in cages **6.9** and **6.10**, respectively, after heating at 70 °C for 18 h (Figure 6.17 a). The ^{1}H NMR spectra of **6.9** and **6.10** were consistent with highly symmetric species of overall *T*-symmetry while ESI-MS confirmed the $Co^{II}_4L_4$ stoichiometry (Figure 6.17 b and c). Single crystals of **6.9** and **6.10** suitable for X-ray diffraction were obtained by Rana Bilbeisi. In both cases, the cages crystallised with approximate *T* point-group symmetry with all Co^{II} stereocentres displaying the same Δ or Λ stereochemistry (Figure 6.17 a). Both cage enantiomers were present in the crystals. Furthermore, the crystal structure of **6.10** contained a central ClO_4^- anion bound within the cavity (Figure 6.21 a). The average Co^{II}-Co^{II} distances for **6.9** and **6.10** were found to be 12.0 Å in both cases which is slightly larger than the previously reported values for the Fe^{II} analogues (11.9 and 11.8 Å respectively).[51]

The distances between the central atom of the ligand and the plane defined by the three adjacent Co^{II} centres varied significantly between **6.9** and **6.10** with values of 1.9 Å in the case of **6.9** and 2.4 Å in the case of **6.10**; evidence of a greater degree of pyramidalisation in the ligand incorporated in **6.10** (Figure 6.18). This phenomenon was already observed in the Fe^{II} analogues of the cages,[51] leading to differences in the cavity size. It was hypothesised that this variation in cavity size could subsequently lead to different host-guest properties.

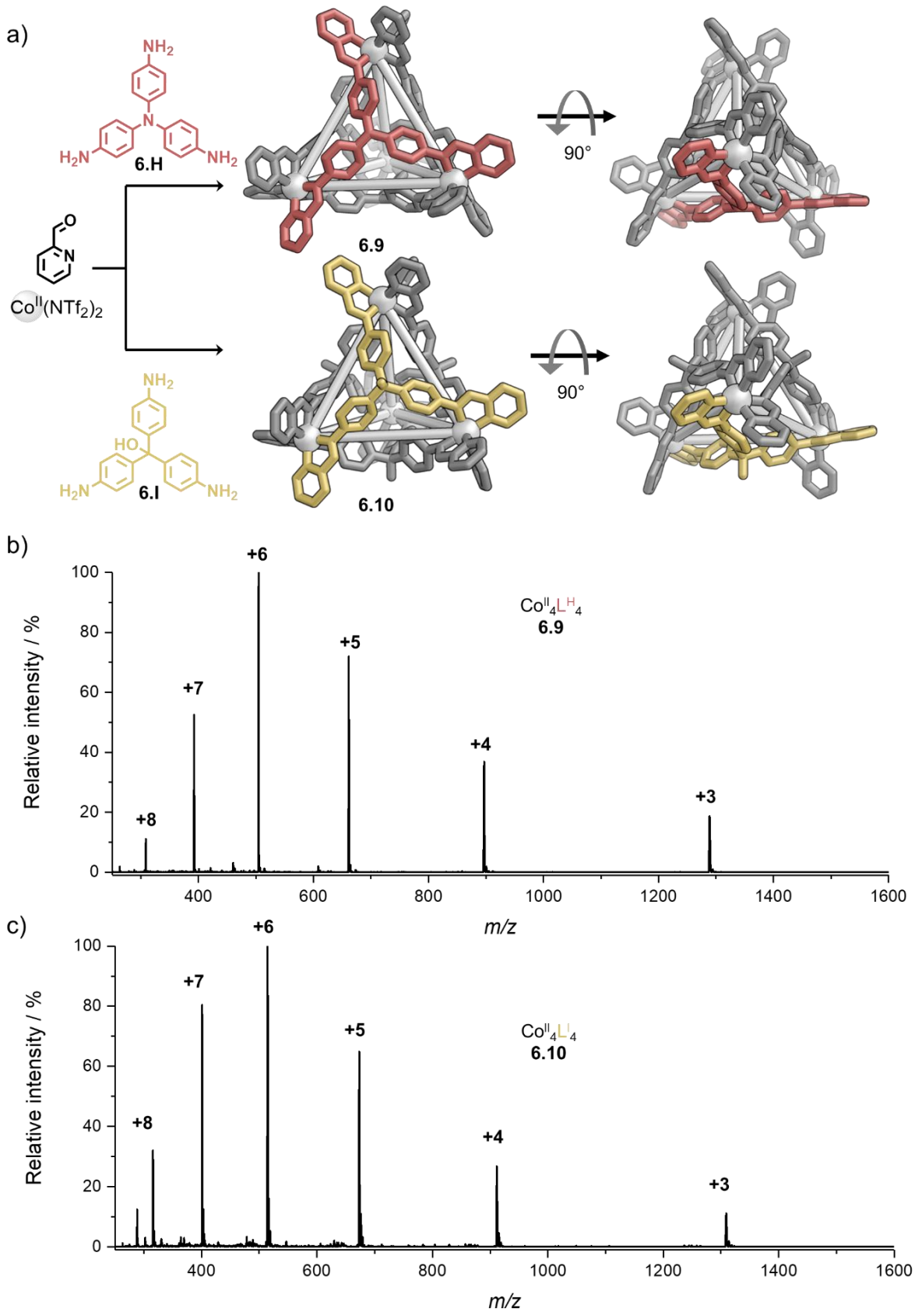
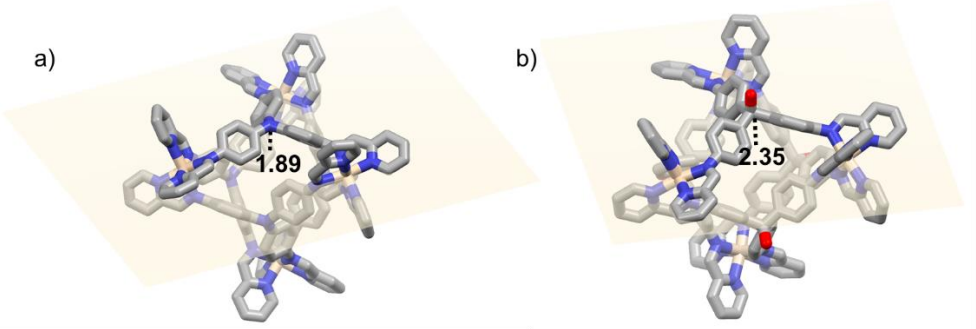


Figure 6.17 | a) Self-assembly of **6.9** and **6.10** from triamine **6.H** or **6.I** respectively, picolinaldehyde and $Co(NTf_2)_2$. Single crystal X-ray structures of **6.9** and **6.10** showing the side view and the top views down the main symmetry axes. Linkages between Co^{II} centres (white spheres) have been added to highlight the framework, hydrogens are omitted for clarity. b and c) ESI-MS spectra of **6.9** and **6.10** respectively.



PLANE	DISTANCE (Å)
1	2.37
2	2.35
3	2.58
4	2.45
AVERAGE	2.44
STD	0.10

Figure 6.18 | Distances between the central N or C of the ligands and the plane (in yellow) defined by three adjacent Co^{II} atoms in the crystal structures of a) **6.9** and b) **6.10**. Table summarizing the distances for **6.10** which were different due to the compound crystallising in monoclinic space group C2/c with no crystallographic symmetry. **6.9** crystallised in cubic space group I222 with crystallographic *T*-symmetry.

6.4.2 Guest binding

After addition of either $TBABF_4$, $TBAClO_4$, TBAI or TBABr, no changes in the ^1H NMR spectra of **6.9** were observed, indicative of no interaction between **6.9** and these anions. The lack of flexibility of **6.H** leading to a smaller cavity size was inferred to prevent encapsulation of anions in **6.9** (Figure 6.19).

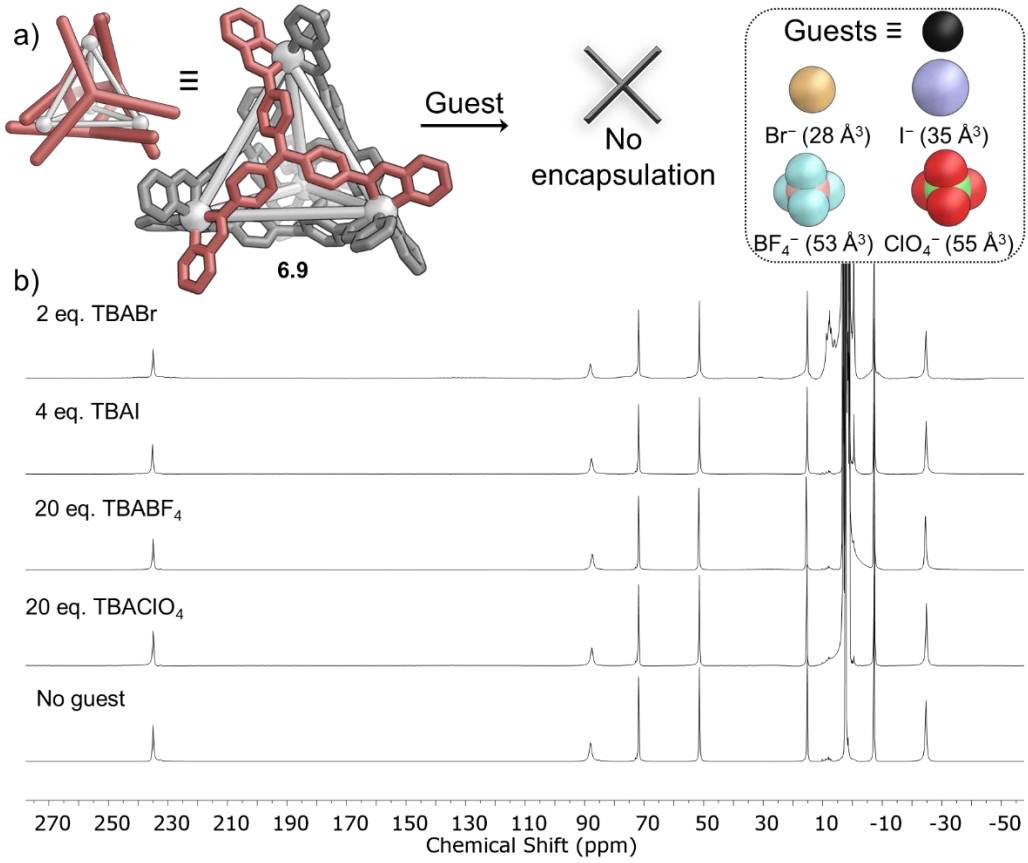


Figure 6.19 | a) Schematic representation of **6.9** and b) Wide sweep ^1H NMR spectra (400 MHz, 298 K, CD_3CN) of **6.9** after addition of TBAI, TBABr, $TBABF_4$ or $TBAClO_4$ (equivalents of guests given on the spectra) showing no changes and hence no anion binding.

In the case of **6.10**, the wide sweep ^{1}H NMR spectrum displayed a second set of signals consistent with a host-guest complex in solution, which was attributed to encapsulated CH_3CN or CD_3CN. The ^{1}H NMR spectrum of **6.10** in CH_3CN was compared to that of the same cage in CD_3CN (Figure 6.20). Solvent suppression (single presaturation at 2.03ppm) was used to observe the region close to the solvent peak, however, due to the paramagnetic character of the sample, this led to a strong decrease in the proton signals of **6.10** (Figure 6.20 b). Both spectra had identical resonances, except for an additional peak at –0.9 ppm in CH_3CN which was attributed to the encapsulated CH_3CN.

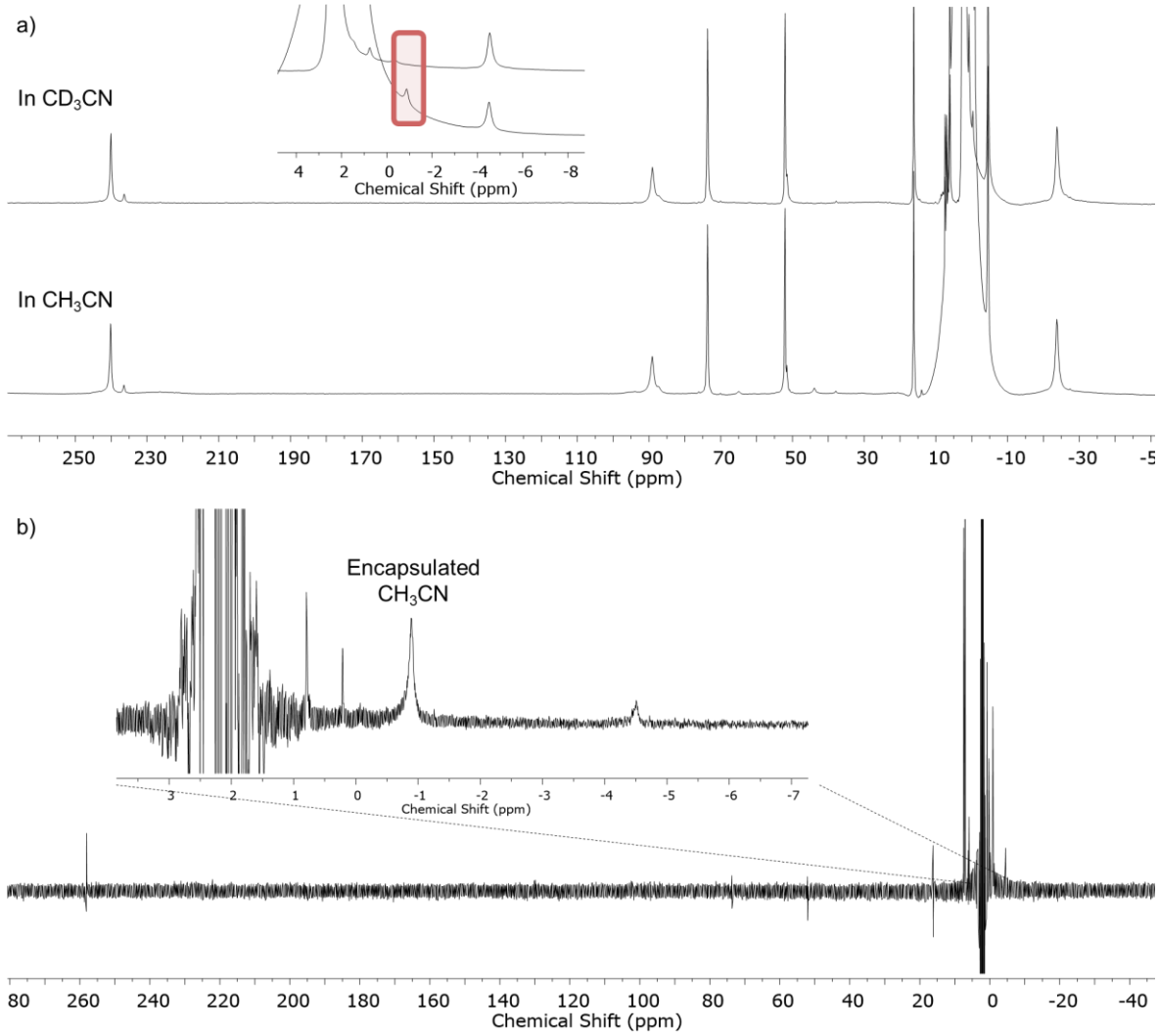


Figure 6.20 | a) Wide sweep ^{1}H NMR spectra (400 MHz, 298 K) of **6.10** in CD_3CN (top) and CH_3CN (bottom) showing the additional peak at 0.9 ppm in the second spectrum. b) Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CH_3CN) after solvent suppression.

Br^-, I^-, BF_4^- and ClO_4^- were bound in slow exchange on the NMR timescale in **6.10**. The subsequent addition of 1 eq. of $TBAClO_4$, $TBABF_4$, TBAI and TBABr to a solution of **6.10** showed that ClO_4^- could be partially displaced by BF_4^-, which was in turn mostly displaced by

I^- and Br^- (Figure 6.21). The association strength of the different anions was inversely correlated to their size, ie. $Br^- > I^- >> BF_4^- > ClO_4^-$.[54]

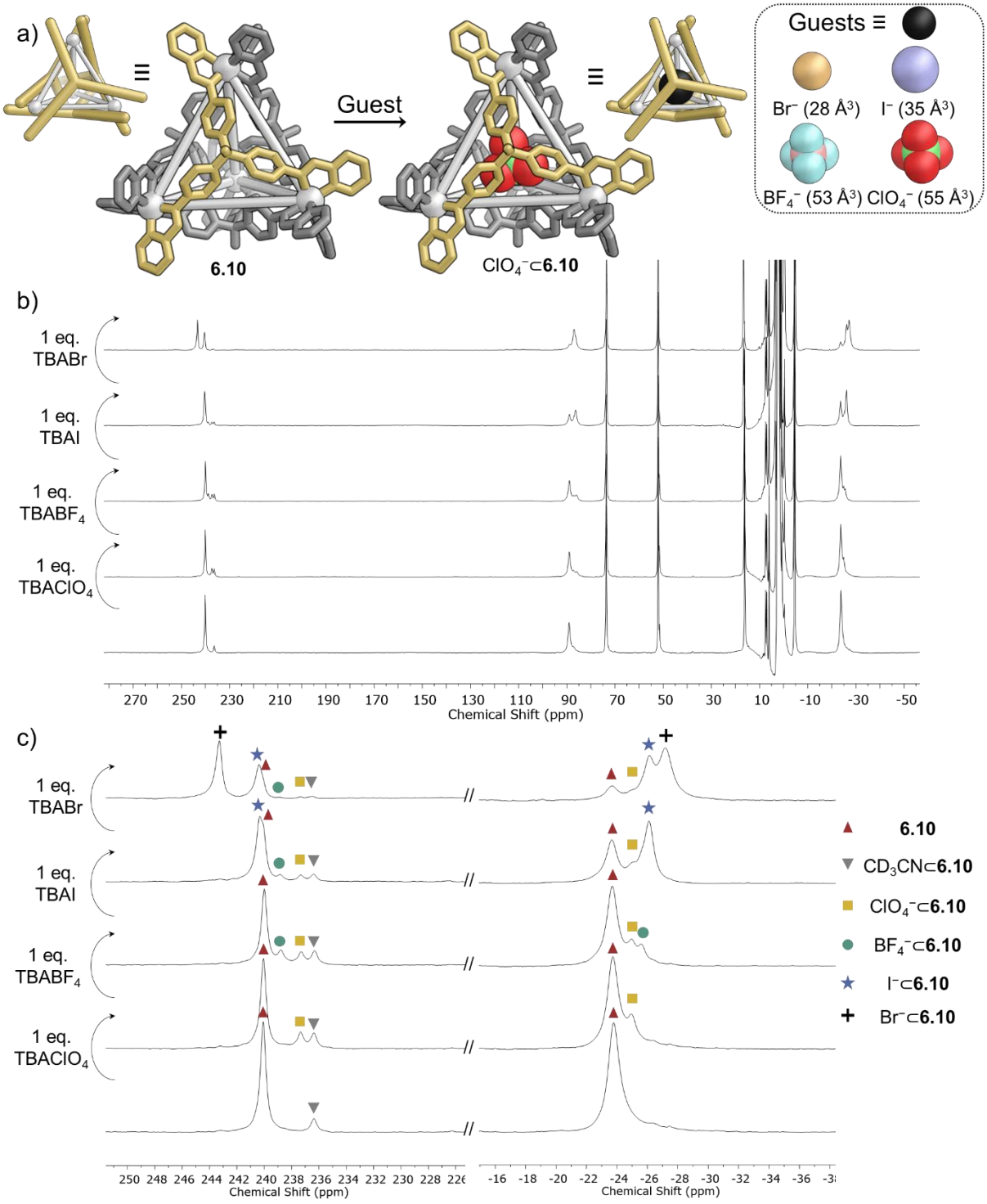


Figure 6.21 | a) Schematic representation of **6.10** and single crystal X-ray structure of **6.10** highlighting the ClO_4^- anion bound within the cavity. b) Wide sweep ^1H NMR spectra (400 MHz, 298 K, CD_3CN) of **6.10** after addition of TBAI, TBABr, $TBABF_4$ or $TBAClO_4$ (1 eq.) showing new sets of peaks for the host-guest complexes. c) Zoom into the 250 to 230 ppm and –16 to –36 ppm region of the ^1H NMR spectra and assignment of the signals for **6.10** (red triangle), $CD_3CN\subset$**6.10** (inverted grey triangle), $ClO_4^-\subset$**6.10** (yellow triangle), $BF_4^-\subset$**6.10** (green circle), $I^-\subset$**6.10** (blue star) and $Br^-\subset$**6.10** (black cross).

The binding of anions within the cavity of **6.10** did not follow the expected 1:1 binding when TBA^+ salts of Br^-, I^-, BF_4^- and ClO_4^- were titrated into a CD_3CN solution of **6.10**. For all anions tested, the concentration of host-guest complex ([HG], $X^-\subset$**6.10**) reached a plateau as the amount of guest added was increased, indicative of competitive binding from another species (Figure 6.23 and Figure 6.24).

Cage **6.10**$(BF_4)_8$ was synthesised directly from $Co(BF_4)_2$, picolinaldehyde and **6.I** and $TBABF_4$ (8 eq.) was added to **6.10**$(NTf_2)_8$, giving two different solutions of **6.10**, containing equal amounts of BF_4^- but different amounts of $TBANTf_2$ (respectively 0 eq. and 8 eq.). Close examination of their ^1H NMR spectra revealed that the amount of encapsulated BF_4^- was greater in the case where neither TBA^+ nor NTf_2^- were present (Figure 6.22 a). Furthermore, when $TBANTf_2$ (50 eq.) was added to **6.10**$(BF_4)_8$, a drastic decrease in the encapsulated BF_4^- was observed (Figure 6.22 b, middle spectrum). When $TBABF_4$ (50 eq.) was added to **6.10**$(BF_4)_8$, the amount of host-guest complex ($BF_4^-\subset$**6.10**) decreased as well, despite the large increase in guest concentration (Figure 6.22, top spectrum). The association of the TBA^+ cation with the anionic guest was thus inferred to limit the formation of the host-guest complex $X^-\subset$**6.10**.

The equations for guest binding were modified to account for the competition of the solvent (CH_3CN) and the TBA^+ in guest binding, with help from John Thoburn. The changes in the concentration of the host-guest complex ($X^-\subset$**6.10**, abbreviated HG), host-solvent complex ($CH_3CN\subset$**6.10**, abbreviated HS), and empty cage (**6.10**, abbreviated H) as a function of the added guest were described by sets of equations. Based on the different behaviours observed in the titrations, two scenario were defined:

- When the affinity of the guest for cage **6.10** was superior to the affinity of the guest for TBA^+, the concentration of host-guest complex [HG] converged to unity and the concentration of host-solvent [HS] converged to zero. This was observed when strongly binding anions (TBABr or TBAI) were titrated into a CD_3CN solution of **6.10**. In this case, the system was described by Equation 6.7 – 6.9.

Equation 6.7

$$[HG] = \frac{[H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}}\right)^2 - 4[H_t][G_t]}}{2}$$

Equation 6.8

$$[H] = \frac{[H_t] - \frac{1}{2}\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}}\right)^2 - 4[H_t][G_t]}\right)}{(1 + K_{HS}[S_t])}$$

Equation 6.9

$$[HS] = \frac{K_S[S_t]}{1 + K_S[S_t]}\left([H_t] - \frac{1}{2}\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}}\right)^2 - 4[H_t][G_t]}\right)\right)$$

- When the affinity of the guest for cage **6.10** was inferior to the affinity of the guest for TBA^+, the concentration of host-guest complex [HG] converged to a value inferior to one and the concentration of host-solvent [HS] remained constant. In this case, the [HS] was fitted to a straight line. A scaling factor (C) was introduced to account for the pairing between TBA and X^-. This was observed when weakly binding anions ($TBABF_4$ or $TBAClO_4$) were titrated into a CD_3CN solution of **6.10**. The system was described by Equation 6.10 – 6.12.

Equation 6.10

$$[HG] = C\frac{[H_t] - [HS] + [G_t] + \frac{1}{K_{HG}} - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}}\right)^2 - 4[H_t][G_t]}}{2}$$

Equation 6.11

$$[H] = [H_t] - [HS] - \frac{C}{2}\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}}\right)^2 - 4[H_t][G_t]}\right)$$

Equation 6.12

$$[HS] = m[G_t] - b$$

where $[S_t]$ represents the total solvent concentration, $[H_t]$ the total host concentration, $[G_t]$ the total guest concentration, K_{HS} the equilibrium constant for solvent binding , K_{HG} the equilibrium

constant for guest binding and m and b are constants. C represented the scaling factor, which was added to account for the contribution of the TBA^+ pairing with the different anions.

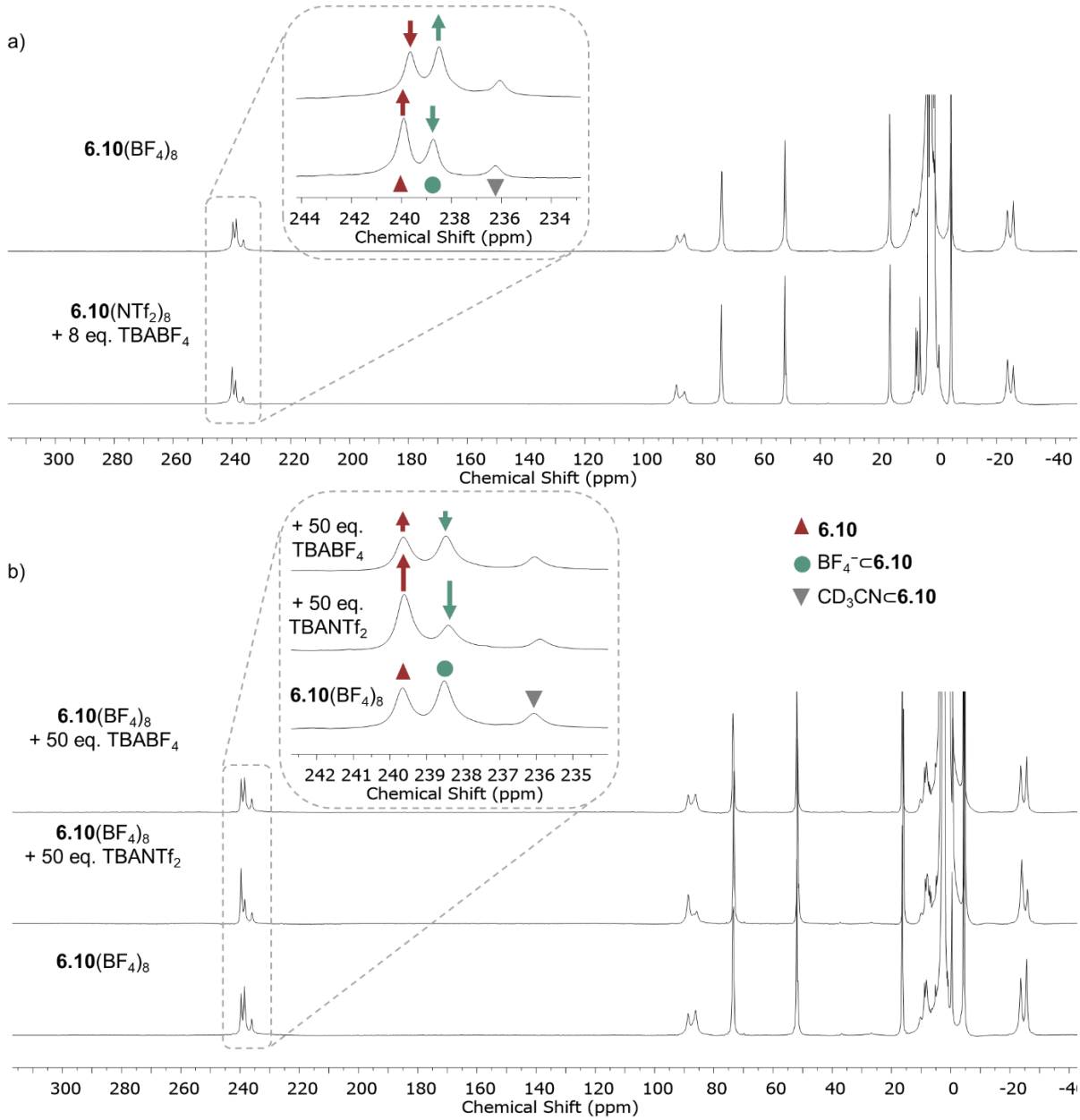


Figure 6.22 | a) Wide sweep ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) of **6.10**$(NTf_2)_8$ to which was added 8 eq. of $TBABF_4$ (bottom) and **6.10**$(BF_4)_8$ (top). b) Wide sweep ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) of **6.10**$(BF_4)_8$ before (bottom) and after addition of either 50 eq. of $TBANTf_2$ (middle) or $TBABF_4$ (top).

Fitting of the NMR titration data to Equation 6.7 – 6.9 or 6.10 – 6.12 simultaneously rather than individually was performed using a Mathematica program written by John Thoburn (Figure 6.23 and Figure 6.24), allowing association constants for the different anions to be calculated. More information on the mathematical models is given in Section 6.6.9.

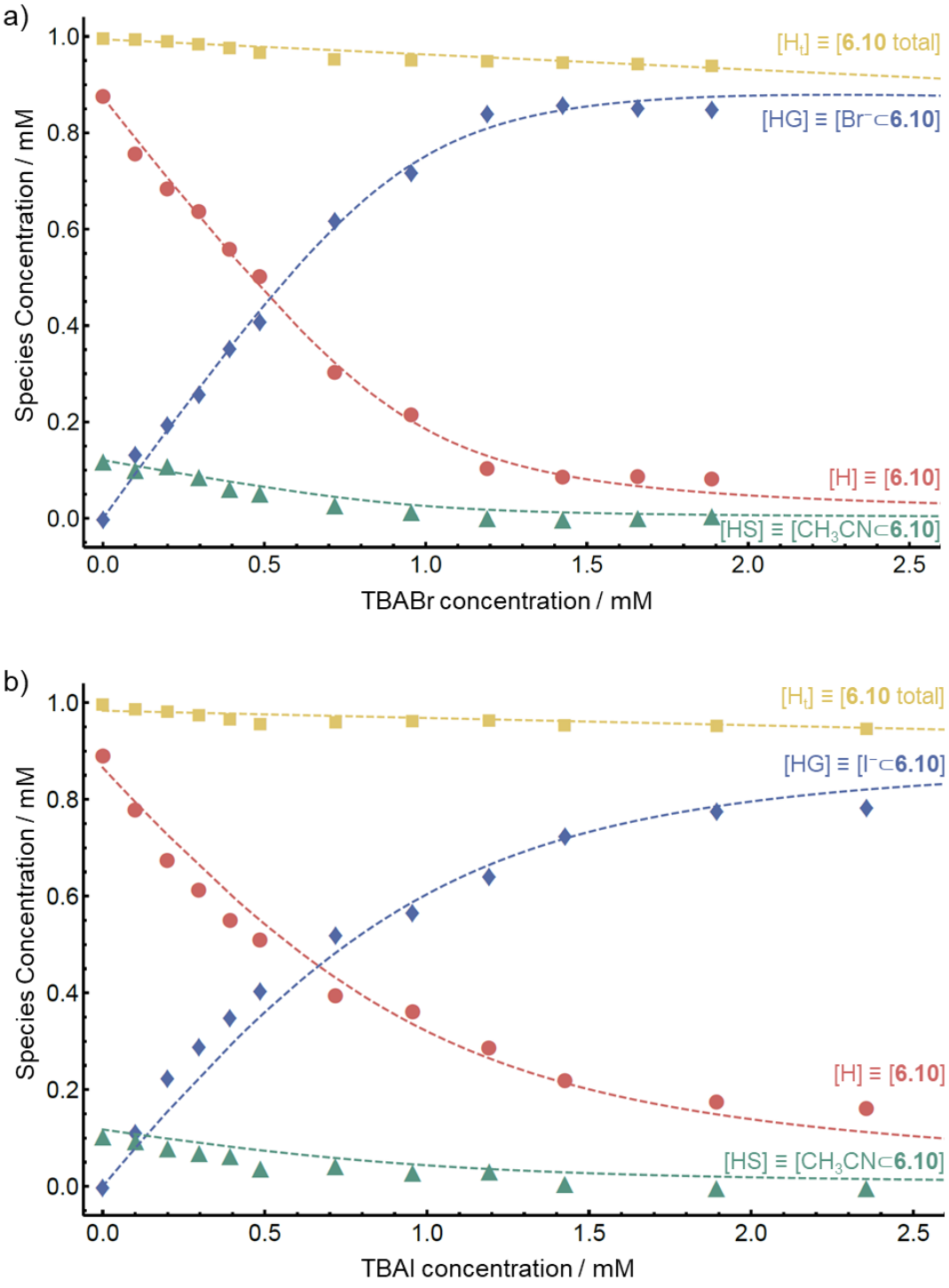


Figure 6.23 | Plot of the concentration of total host (**6.10** total in yellow), free host (**6.10** free in red), host-guest complex (X^-⊂**6.10** in blue) and solvent-host complex (CH_3CN⊂**6.10** in green) as a function of the concentration of guest (TBAX) for a) X = Br^- and b) I^-. The experimental data obtained from titrations is given by the points (triangles, circles, squares and diamonds) and the fitting obtained is represented by the dotted lines.

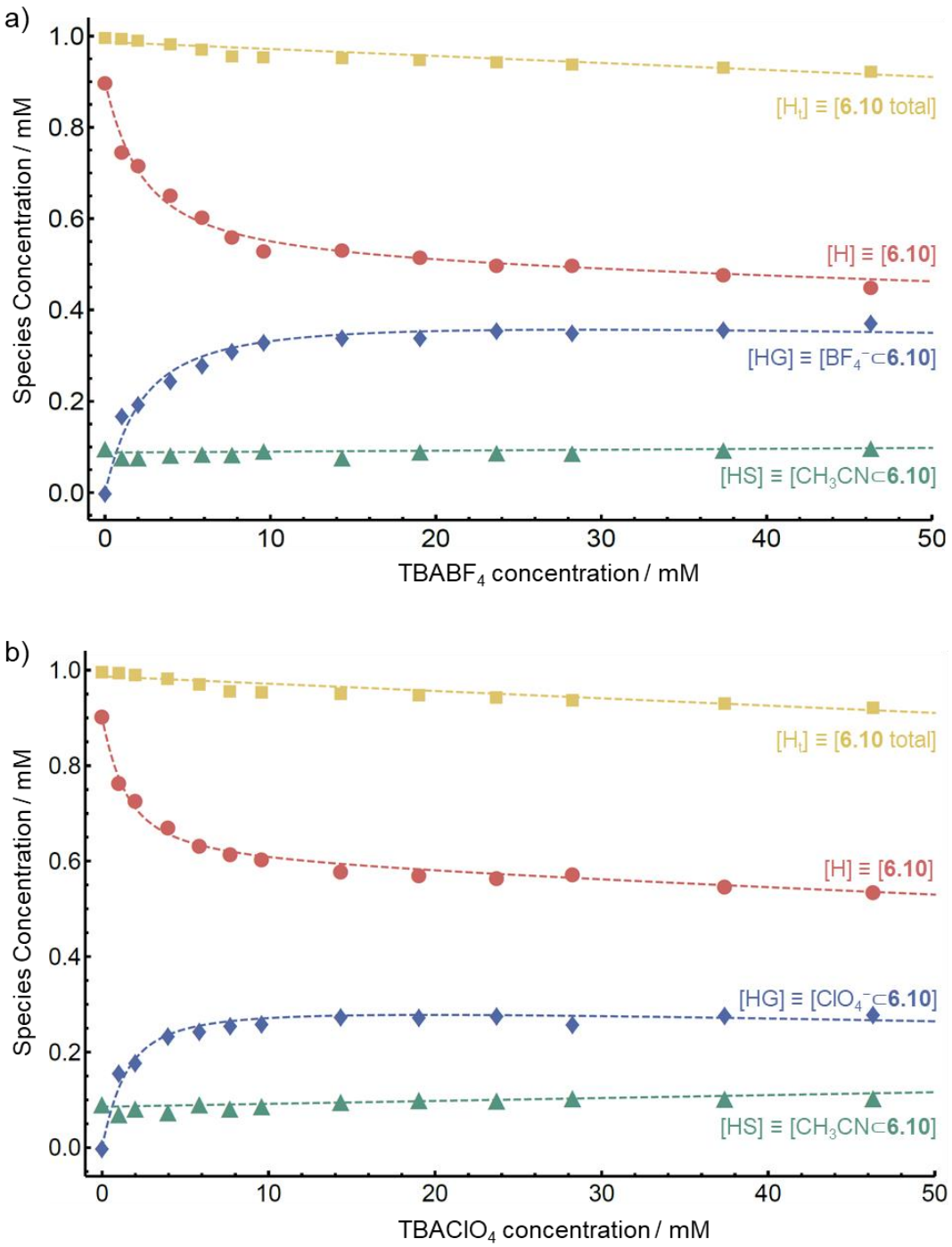


Figure 6.24 | Plot of the concentration of total host (**6.10** total in yellow), free host (**6.10** free in red), host-guest complex (X^-⊂**6.10** in blue) and solvent-host complex (CH_3CN⊂**6.10** in green) as a function of the concentration of guest (TBAX) for a) X = BF_4^- and b) ClO_4^-. The experimental data obtained from titrations is given by the points (triangles, circles, squares and diamonds) and the fitting obtained is represented by the dotted lines.

Due to dilution effects resulting from the addition of the guest solution during titration, the total host concentration [H_t] decreased and was modelled as a straight line (Figure 6.23 and Figure 6.24, yellow line). Thus, the host-guest concentration [HG] increased asymptotically to the dilution curve (Figure 6.23 and Figure 6.24, blue line). Furthermore, unity was not reached due

to the competition of TBA^+ for the guest, which was introduced into the system in the same amounts as the guest. The scaling factor C (Equation 6.7 – 6.9) was thus introduced in the equations to account for the TBA pairing; C was equal to one in the case of Br^- and I^- and corresponded to the saturation concentration, ie. the asymptote of the [HG] curve for BF_4^- and ClO_4^-. For large guests (BF_4^- and ClO_4^-), the concentration of encapsulated solvent [HS], did not decrease asymptotically to zero as more guest was added, but remained at 10% occupancy and the binding curves for [HS] were thus fitted to a straight line. This was explained by the weaker binding affinity of these guests which could not displace the solvent, present in large excess. Smaller guests (Br^- and I^-) with stronger binding could expel the weakly bound solvent.

The following values for the association constants of the guests in **6.10** in the presence of TBA^+ were obtained:

$$K_{Br} = 16000 \pm 1000\ M^{-1}$$

$$K_{I} = 4800 \pm 400\ M^{-1}$$

$$K_{BF4} = 580 \pm 50\ M^{-1}$$

$$K_{ClO4} = 1100 \pm 100\ M^{-1}$$

$$K_{CH_3CN} = 0.0072 \pm 0.0004\ M^{-1}$$

These values take into account the non-specific interaction of TBA^+ and the anion and thus do not represent the actual binding affinities of the anion within **6.10**. The contribution of each phenomenon individually (internal binding and ion pairing) could not be quantified as the system was underdetermined.

6.4.3 Quantification by mass spectrometry

The similarity in sizes of **6.H** and **6.I** has been used previously to generate a family of heteroleptic cages $Fe^{II}L^{H}_{x}L^{I}_{4-x}$.[51] The substitution of Fe^{II} for Co^{II} leading to guest encapsulation in **6.10** opens prospects for selectively templating specific members of the library. As expected, the reaction of **6.H** and **6.I** (1 eq. each), picolinaldehyde (6 eq.) and $Co(NTf_2)_2$ (2 eq.) in CH_3CN led to the formation of a library (**Lib$_{NTf_2}$**) of cages of formula $Co^{II}_4L^{H}_{x}L^{I}_{(4-x)}$ as confirmed by ESI-MS (Figure 6.25).

The same reaction in the presence of sub-stoichiometric amounts of picolinaldehyde (3 eq. instead of 6 eq.) at room temperature led to the formation of a different library (**Lib$_{NTf_2SUB}$**). ESI-MS revealed the presence of congeners of stoichiometry $Co^{II}_4L^{H}_{x}L^{I}_{(4-x)}$ with $2 \leq x \leq 4$

predominantly (Figure 6.26 a). Three sets of signals containing six, four, and one resonance each were observed in the imine region of the ^{1}H NMR (Figure 6.26 b) which were attributed to $Co^{II}_4L^{H}_4$ (*T* symmetry, one signal per imine proton), $Co^{II}_4L^{H}L^{I}_3$ (C_3 symmetric, four signals) and $Co^{II}_4L^{H}_2L^{I}_2$ (C_2 symmetric, six signals). Based on this assignment, each signal observed in the imine region of the ^{1}H NMR spectrum of **Lib$_{NTf2}$** was attributed to one of the congeners present (Figure 6.26 c). However, the strong overlap of some signals limited the possibility of signal deconvolution, thus no information on the proportion of species present in solution was obtained.

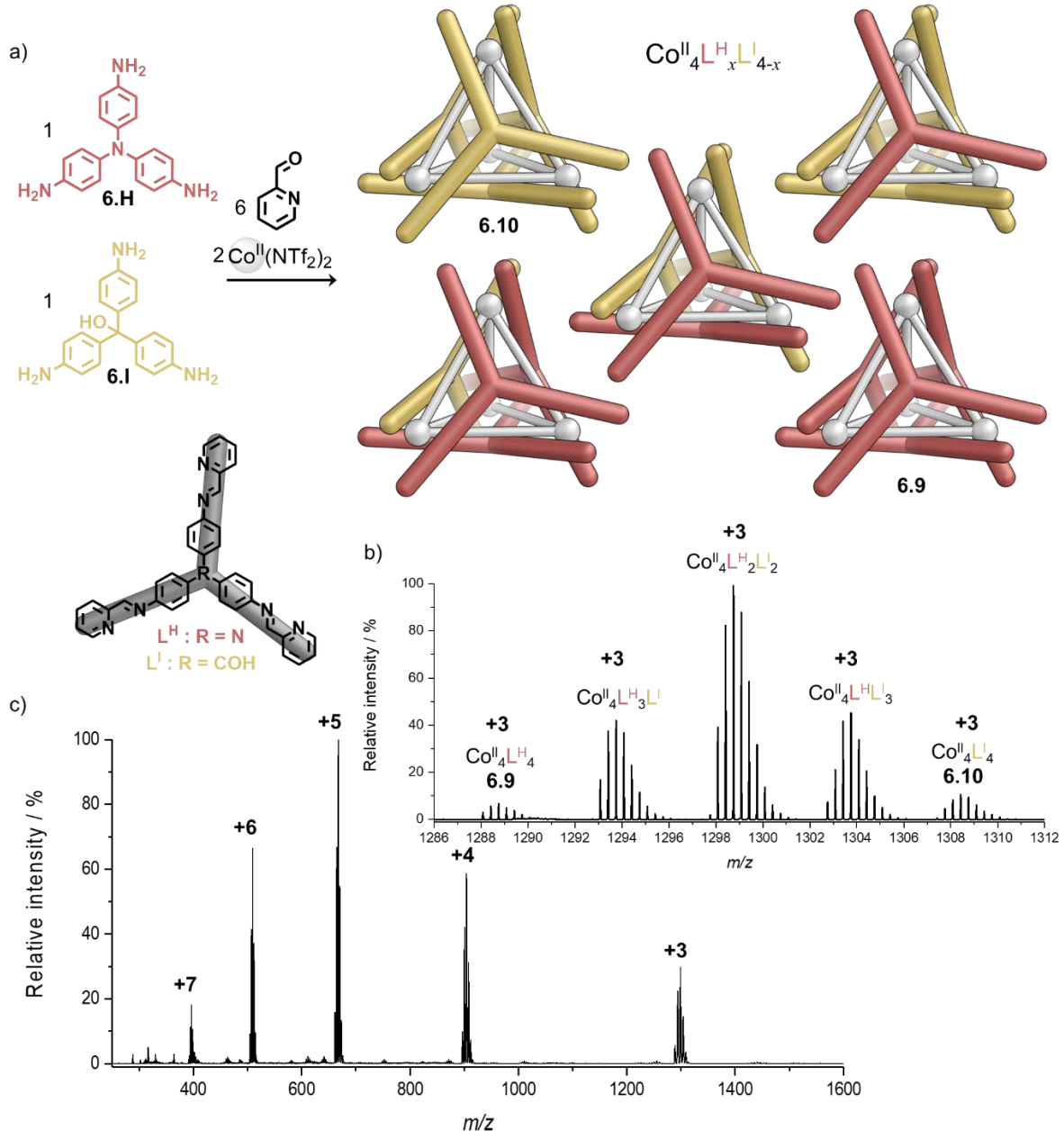


Figure 6.25 | a) Self-assembly of **Lib$_{NTf2}$** from **6.I**, **6.H**, picolinaldehyde and Co(NTf$_2$)$_2$ into a library of cages of the formulae $Co^{II}_6L^{H}_xL^{I}_{4-x}$. b) +3 charge peak region of the HR-ESI-MS and c) LR-ESI-MS of **Lib$_{NTf2}$** showing the isotopic pattern for the cages $Co^{II}_6L^{H}_xL^{I}_{4-x}$.

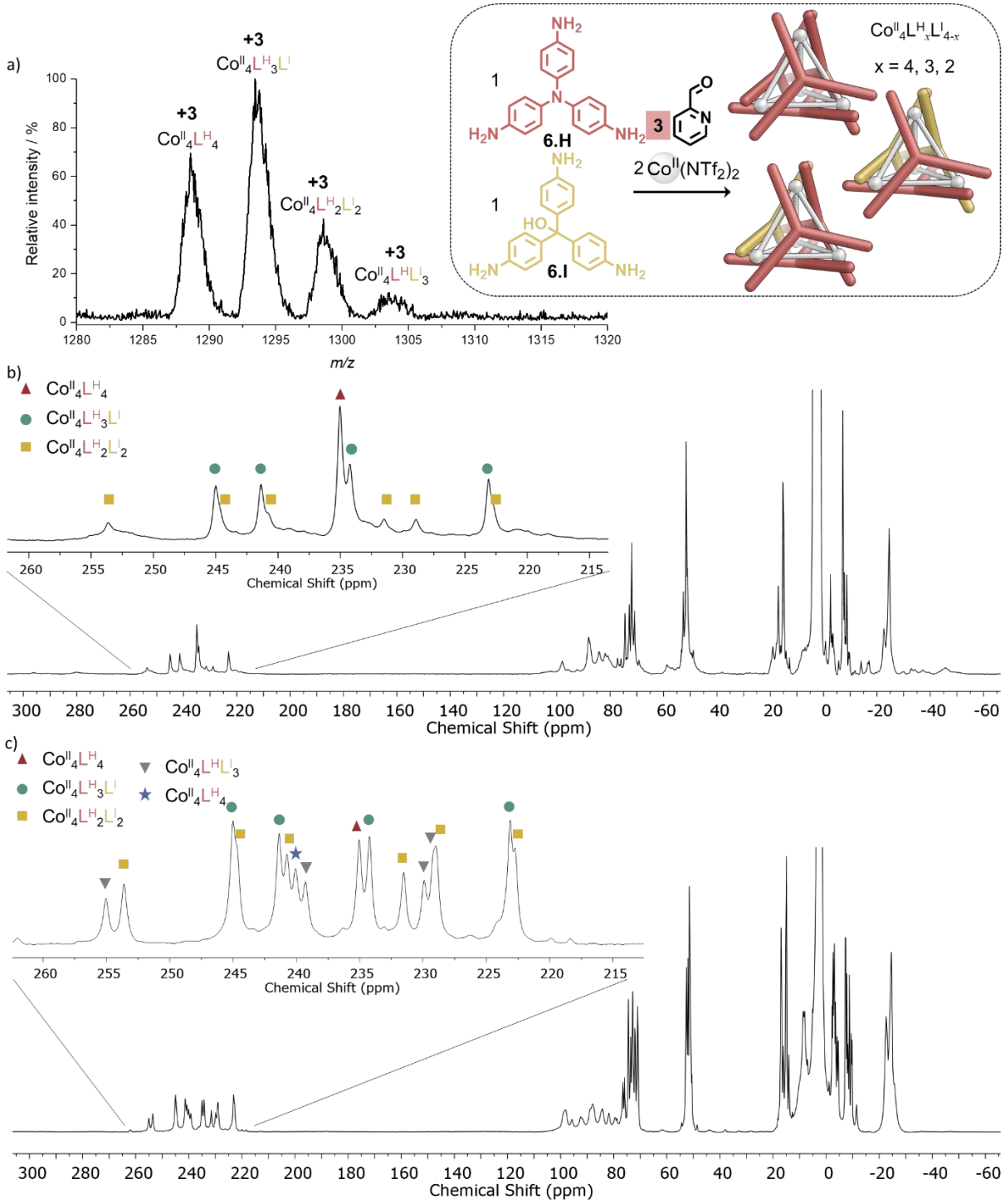
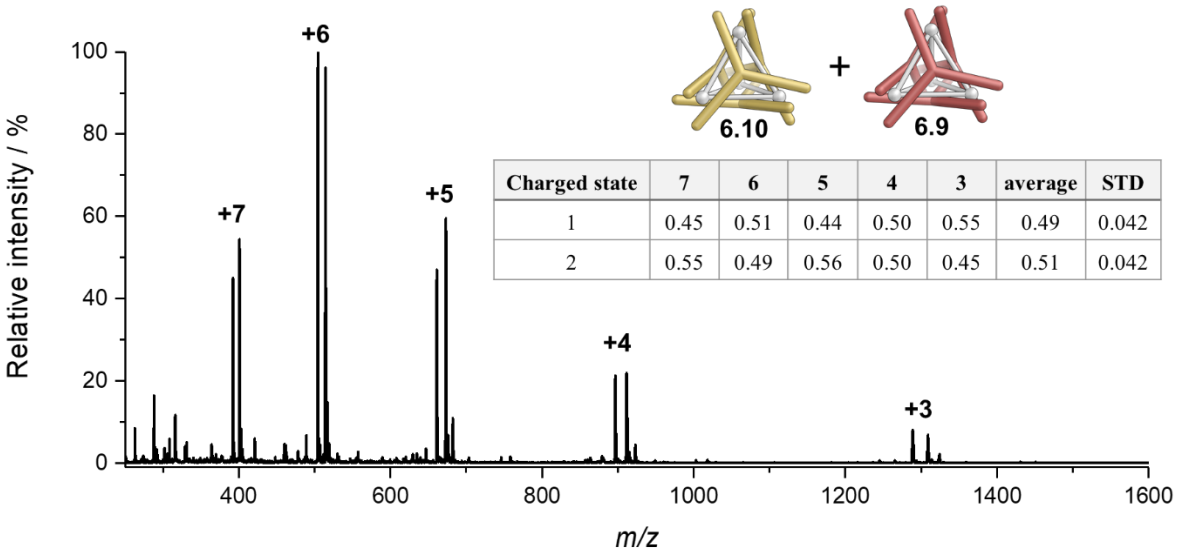


Figure 6.26 | a) LR-ESI-MS of **Lib$_{NTf2SUB}$** showing that the predominant species are those incorporating high numbers of L^{H}. Wide sweep ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) of b) **Lib$_{NTf2SUB}$** with the signals assigned based on the symmetry of each species and of c) **Lib$_{NTf2}$** with all signals attributed in the imine region. Signals for Co$^{II}_4$L$^{H}_4$ are marked by a red triangle, Co$^{II}_4$L^{H}L$^{I}_3$ by a green circle, Co$^{II}_4$L$^{H}_2$L$^{I}_2$ by a yellow triangle, Co$^{II}_4$L$^{H}_3$L^{I} by a grey inverted triangle and Co$^{II}_4$L$^{I}_4$ by a blue star.

Mass spectrometry, in contrast, allowed the library speciation to be studied in greater detail. When **6.9** and **6.10** were combined in equimolar amounts and quickly analysed by ESI-MS

before scrambling of the ligands could occur, similar ESI response factors were observed for both cages (Figure 6.27). The relative integrals of the peaks for each species did not deviate significantly from the expected 1:1 ratio when averaged across the observable charge states of +3 to +7. It was therefore hypothesised that the heteroleptic congeners of the library have similar response factors and that the integrals of the signals observed for each species are proportional to the species concentration in solution.



Charged state	7	6	5	4	3	average	STD
1	0.45	0.51	0.44	0.50	0.55	0.49	0.042
2	0.55	0.49	0.56	0.50	0.45	0.51	0.042

Figure 6.27 | LR-ESI-MS of a freshly combined equimolar mixture of **6.9** and **6.10** and the table summarising the integrals of each species for the observed charge states.

In the light of this result, mass spectrometry was used to quantify the amount of congeners within libraries formed from **6.H** and **6.I**. Within each charge state observed in the mass spectra, the integrals of the *m/z* peaks were normalised and the values obtained were averaged across the +7 to +3 charge states. Experiments were repeated three times to minimise the error.

In the case of $\mathbf{Lib_{NTf_2}}$ (Figure 6.25), the $Co^{II}_4L^H_xL^I_{(4-x)}$ species were observed in a 1:4.1:5.8:3.8:0.8 ratio (Figure 6.25 b). Similarly to Section 6.1.2, a set of equilibrium constants between congeners could be calculated. These equilibrium constants were defined by the ratio between the observed amount of a species in the library and a binomial distribution. Each congener was identified by the number *N* of ligand L^H they possess. The relative energy of each could be determined by using the same formula as previously:

Equation 6.1

$$\Delta E_{rN} = -RTln(K_N), \quad with\ K_N = \frac{\left(\frac{I_N}{I_{statN}}\right)}{I_0}$$

where I_N represented the normalised integrals of the mass peak for the structure incorporating *N* L^H, I_{statN} was the corresponding normalised integral for an ideal binomial distribution, and I_0

was the normalised integral of the mass peak for the most stable structure, here **6.10** (N = 0). The average values of the relative energies of each congener were obtained by averaging the values across all charge states observed and across three repeats of each experiment. The relative Gibbs energies showed no significant deviation from the null hypothesis by Fisher's test as the p-value was above the significance threshold of 0.05 (F = 0.262 and p = 0.660). The distribution of congeners within $\mathbf{Lib_{NTf_2}}$ was thus inferred to not significantly deviate from the binomial distribution (Figure 6.33, grey curve).

Addition of anions X^- (2 eq., Br^-, I^-, BF_4^- or ClO_4^-) able to bind in the cavity of **6.10** to $\mathbf{Lib_{NTf_2}}$ followed by equilibration of the mixtures for 18 h at 70 °C led to changes in the proportion of species present in the library ($\mathbf{Lib_X}$). The ESI-MS spectra displayed clusters of peaks indicative of the $Co^{II}_4L^H_xL^I_{(4-x)}$ cages associated with zero, one or more X^- (Figure 6.28 to Figure 6.31). The presence of a signal for cages with no X^- associated ($Co^{II}_4L^H_xL^I_{4-x}$ + $mNTf_2$) along with signals for **6.9** (which doesn't bind X^-) associated with X^- indicated that X^- can either be associated externally or internally with the cages. The species within the library $\mathbf{Lib_X}$ which do not have an encapsulated X^- were defined as belonging to a distribution named **DisA**(X^-, n) with n the number of anions X^- externally associated ($0 \leq n < 8$). The species within $\mathbf{Lib_X}$ which do have an encapsulated X^- were defined as belonging to a distribution named **DisB**(X^-, n) with n the overall number of anions X^- associated (internally and externally, $1 \leq n < 8$). The new library $\mathbf{Lib_X}$ thus consists of a distribution of species with:

- no X^- associated ($[Co^{II}_4L^H_xL^I_{4-x}$ + $mNTf_2]$, $0 < m < 8$) which is referred to as **DisA**(X^-,0))
- X^- externally associated ($[Co^{II}_4L^H_xL^I_{4-x}$ + nX^- + $(m–n)NTf_2]$, $1 \leq n < 8$, $n < m < 8$) which is referred to as **DisA**(X^-,n)
- X^- internally associated ($[X^- \subset Co^{II}_4L^H_xL^I_{4-x}$ + $(n–1)X^-$ + $(m–n)NTf_2]$, $1 \leq n < 8$, $n < m < 8$) which is referred to as **DisB**(X^-,n))

Information on the impact of anion X^- on the library can be obtained by analyzing the distribution of congeners $X^- \subset Co^{II}_4L^H_xL^I_{4-x}$ of the library $\mathbf{Lib_X}$ which are represented by **DisB**(X^-,n). However, the peaks corresponding to **DisB**(X^-,n) overlap with those of **DisA**(X^-,n) in the mass spectra as they represent complexes of the same *m/z* ratios (Figure 6.28 to Figure 6.31, blue triangle and grey star). The integrals of the *m/z* peaks in the ESI-MS were calculated for each congener within each charge state observed for the cluster with zero and one anion associated (Figure 6.28 to Figure 6.31, blue triangle and green circle). Due to similarity of the response factors for **6.9** and **6.10** across charge states, the integrals of the *m/z* peaks were assumed to be proportional to the concentration of each species in solution.

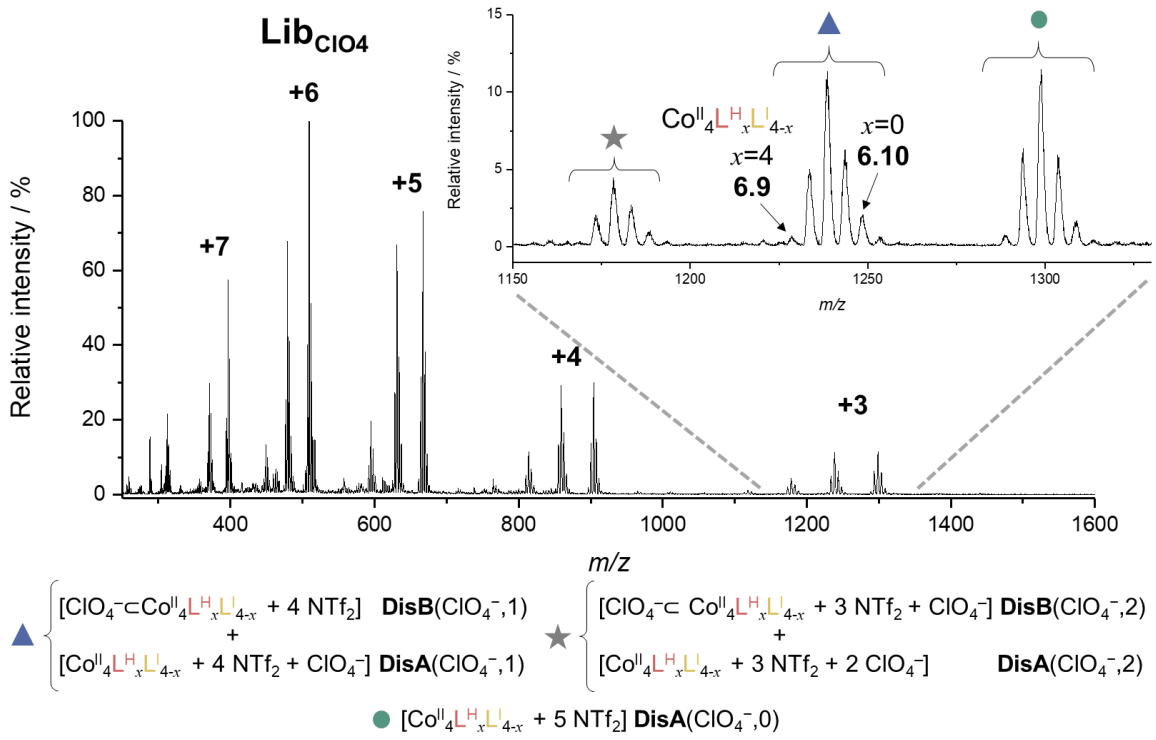


Figure 6.28 | LR-ESI-MS of **Lib$_{ClO4}$** obtained after addition of TBAClO$_4$ (2 eq.) to **Lib$_{NTf2}$**. Zoom in the +3 region of the mass spectra showing the clusters corresponding to cages with no ClO$_4^-$ associated (**DisA**(ClO$_4^-$,0), green circle), cages with one ClO$_4^-$ associated (**DisA**(ClO$_4^-$,1) + **DisB**(ClO$_4^-$,1), blue triangle) and cages with two ClO$_4^-$ associated (**DisA**(ClO$_4^-$,2) + **DisB**(ClO$_4^-$,2), grey star).

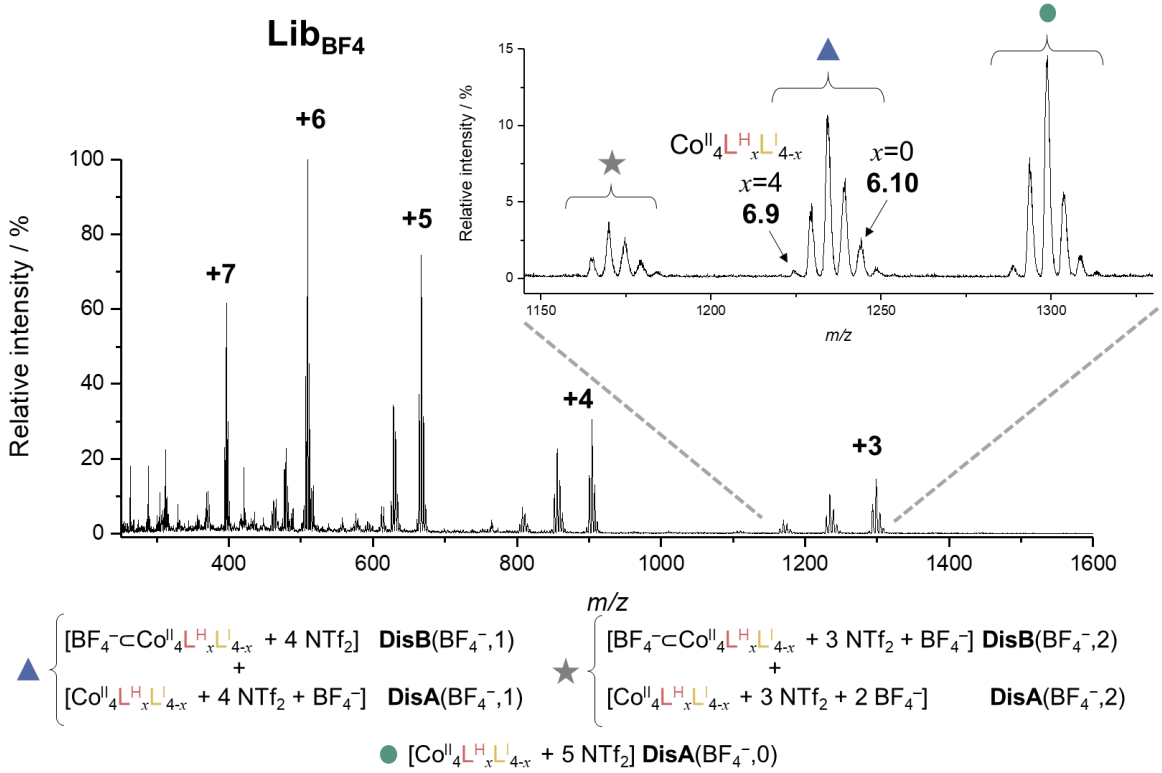


Figure 6.29 | LR-ESI-MS of **Lib$_{BF4}$** obtained after addition of TBABF$_4$ (2 eq.) to **Lib$_{NTf2}$**. Zoom in the +3 region of the mass spectra showing the clusters corresponding to cages with no BF$_4^-$ associated (**DisA**(BF$_4^-$,0), green circle), cages with one BF$_4^-$ associated (**DisA**(BF$_4^-$,1) + **DisB**(BF$_4^-$,1), blue triangle) and cages with two BF$_4^-$ associated (**DisA**(BF$_4^-$,2) + **DisB**(BF$_4^-$,2), grey star).

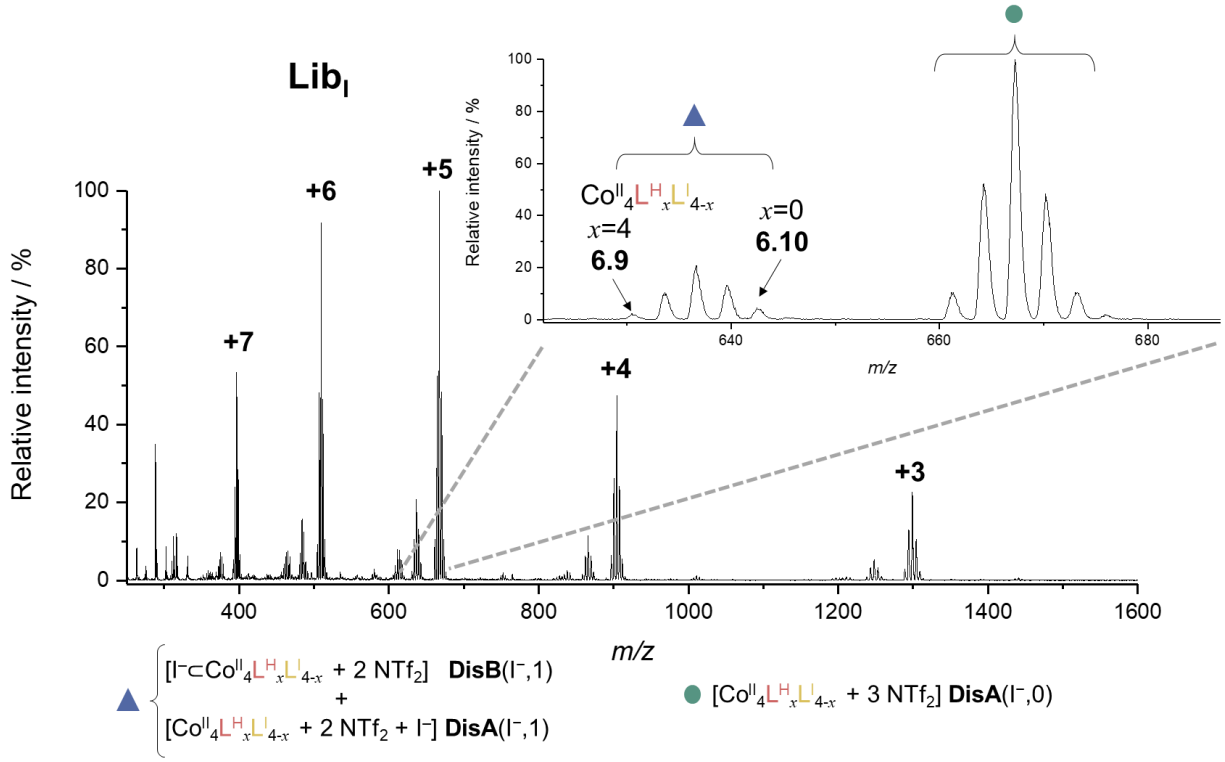


Figure 6.30 | LR-ESI-MS of **Lib$_I$** obtained after addition of TBAI (2 eq.) to **Lib$_{NTf2}$**. Zoom in the +3 region of the mass spectra showing the clusters corresponding to cages with no I^- associated (**DisA**(I^-,0), green circle) and cages with no I^- associated (**DisA**(I^-,1) + **DisB**(I^-,1), blue triangle).

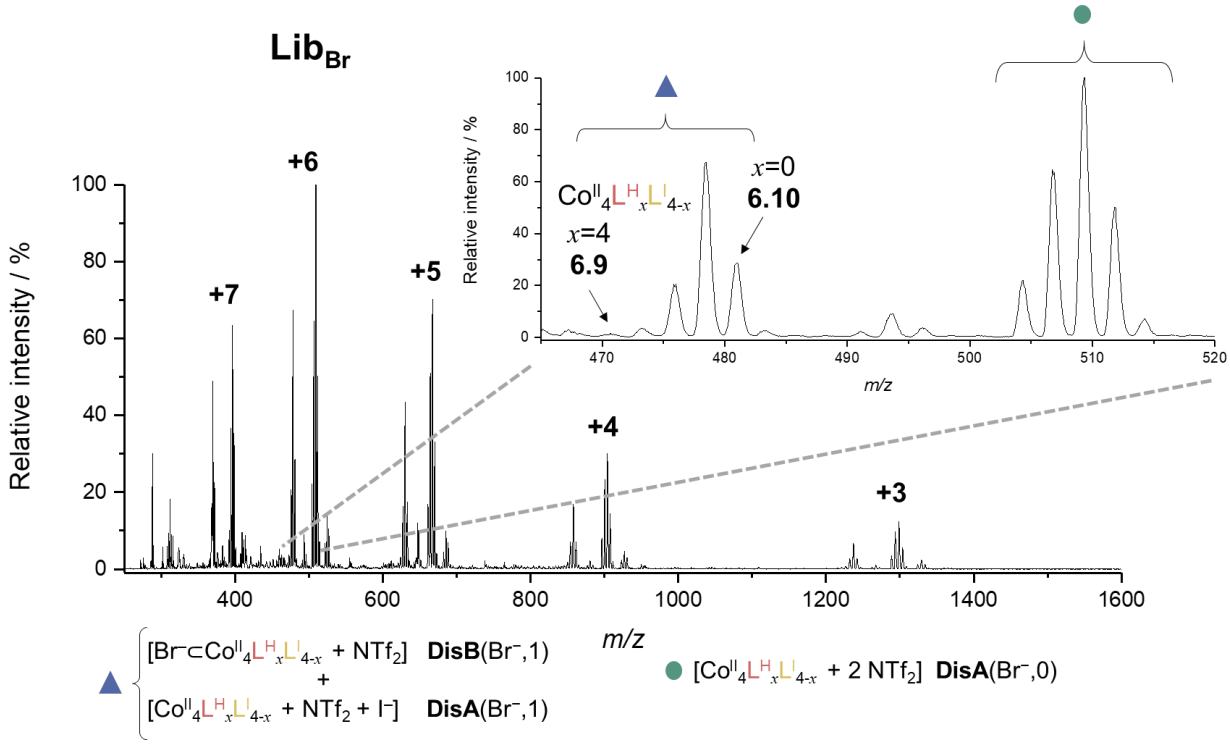


Figure 6.31 | LR-ESI-MS of **Lib$_{Br}$** obtained after addition of TBABr (2 eq.) to **Lib$_{NTf2}$**. Zoom in the +3 region of the mass spectra showing the clusters corresponding to cages with no Br^- associated (**DisA**(Br^-,0), green circle) and cages with one Br^- associated (**DisA**(Br^-,1) + **DisB**(Br^-,1), blue triangle).

DisA(X^-,1) is made of the same cages, as **DisA**(X^-,0) (empty cages $Co^{II}_4L^H_xL^I_{4-x}$) and only differ by the number of anions X^- externally associated. The number of externally associated anions observed is only dependent on the ionisation of the mass spectrometer. Therefore the ratio of congeners within each clusters are identical and the distributions only differ by a scaling factor. As **6.9** can not encapsulate any anions (as observed in the NMR, Figure 6.19), the peak corresponding to this cage (marked by x=4 on the blue triangle cluster in Figure 6.28 to Figure 6.31) is representative of **DisA**(X^-,1) exclusively. The scaling factor between **DisA**(X^-,1) and **DisA**(X^-,0) could thus be obtained by comparing the integrals of the peak for **6.9** in the two adjacent cluster (Figure 6.28 to Figure 6.31, blue triangle and green circle). It was thus possible to obtain the values of the integrals for the congeners within **DisB**(X^-,1) by subtracting **DisA**(X^-,1) from the values observed in the cluster incorporating both of these distributions (Figure 6.28 to Figure 6.31, blue triangle).

The values obtained for the congeners within **DisB**(X^-,1) were normalized to unity and averaged across all charge states observed. The distributions obtained deviated strongly from the near binomial distribution observed for **Lib$_{NTf2}$** (Figure 6.32). Structures incorporating high numbers of L^I were favoured for all anions tested, with greater deviations for smaller anions. This correlated with the trends observed for binding of the anions in **6.10**, ie. $Br^- > I^- >> BF_4^- > ClO_4$.

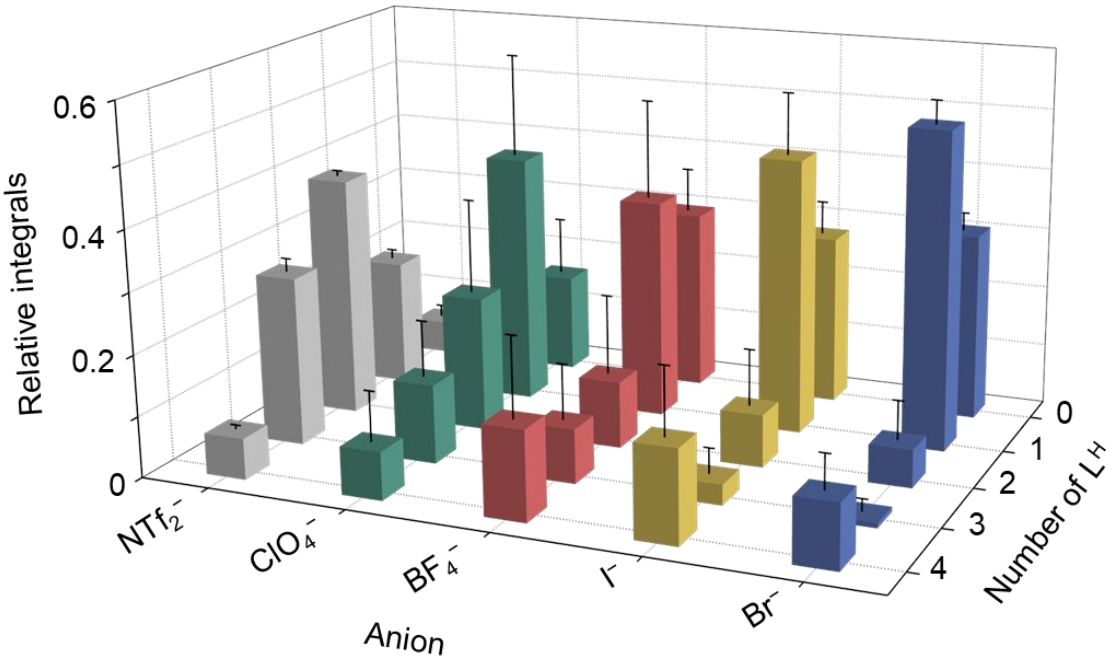


Figure 6.32 | Normalised integrals for **Lib$_{NT2}$** (gray) and **DisB**(X^-) obtained from **Lib$_X$** (X^- = Br^-, I^-, BF_4^- and ClO_4^-, respectively in blue, yellow, red and green). Average and standard deviations (error bars) over three repeats of the experiment and across all charge states observed.

The Gibbs energies were calculated for the congeners of $\mathbf{Lib_X}$ ($X^- = Br^-$, I^-, BF_4^- and ClO_4^-) using Equation 6.1 and the ratio between the integrals observed for **DisB**(X^-) and those observed for the library in the absence of encapsulated anions ($\mathbf{Lib_{NTf2}}$) was determined. The values of the relative energies were plotted as a function of the number of L^H incorporated in the congeners for each anion used. The energetic cost was higher for each L^H incorporated, due to the reduced size of the cavity of the congeners and hence the decreased binding affinity for the anions. Anions which bound strongly within **6.10** imparted greater changes on the libraries, leading to higher levels of energetic destabilisation of structures incorporating L^H predominantly. The relative energy calculated for **6.9** was lower than for the $Co^{II}_4L^H_3L^I$ congener which was explained by the excess of L^I left in the mixture after amplification of the congeners incorporating more L^H (Figure 6.33). In this case, the ESI-MS analysis enabled information to be gained on the binding preference of anions within a library of compounds.

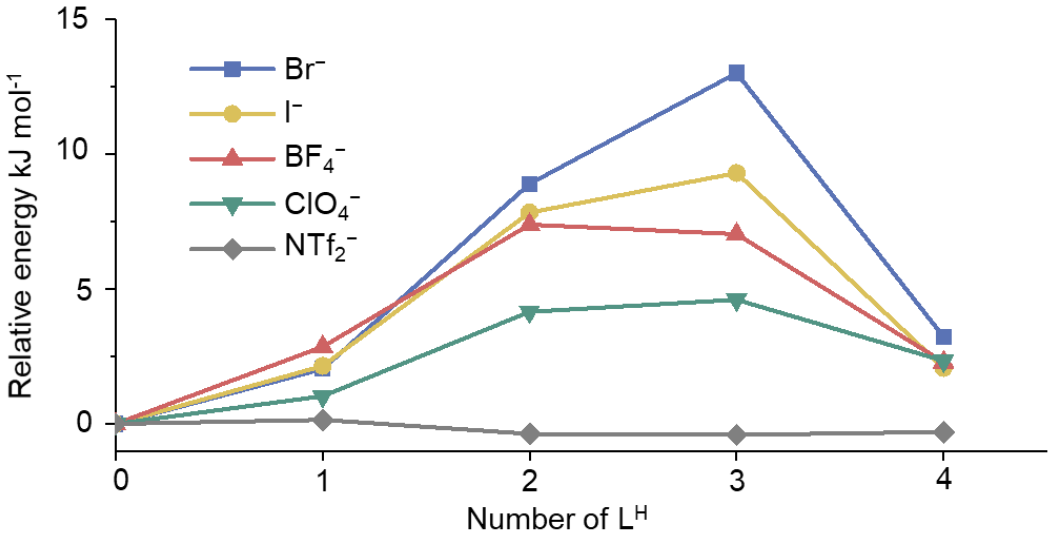


Figure 6.33 | Relative energies (kJ mol^{-1}) of the library of cages in the absence of templating anions (NTf_2^-, grey) and of the libraries of cages in the presence of templating anions (Br^-, I^-, BF_4^- and ClO_4^- respectively in blue, yellow, red and green). Average over three repeats of the experiment and across all charge states observed.

6.5 Conclusion and future work

In the first example of this chapter, a straightforward analysis based on ESI-MS was used to elucidate some of the key factors that led a fluorinated bis-bidentate ligand to form *mer* rather than *fac* metal vertices. The energetic preference of fluorinated ligands to form *mer* vertices, and therefore prismatic structures, was quantified and compared to the same value for non-fluorinated ligands. The relative energetic destabilisation was found to be greater for fluorinated ligands incorporated into *fac* vertices and thus tetrahedra than for non-fluorinated ligands incorporated into *mer* vertices and thus prisms. With this key information in hand, the design of more complex, larger supramolecular architectures with *mer*-coordination motifs such as icosahedra or dodecahedra could be facilitated.

In the second case, fine-grained details of self-sorting within complex cage systems could be quantified by various MS techniques. A highly accessible analysis based on ESI-MS was used to deduce the quantity and composition of metal-organic cages within self-sorted mixtures. In addition, the IM-MS study enabled the morphology of heteroleptic species to be determined. Combined, these two techniques reported on the distribution of products within self-assembled mixtures and the sorting regime they adopted. Rules governing the self-assembly of heteroleptic structures based on ligand properties were drawn, which could help with building and understanding more complex systems in the future.

Finally, the subtle effects of guest encapsulation on a statistical library of cages was studied by ESI-MS using a similar method as in the first example. The presence of the anion favoured structures incorporating more of the flexible ligand. Strongly bound anions lead to an increased energetic destabilisation of each congener of the library compared to weaker binding anions, allowing for the information on relative binding strength to be gained *via* MS. Quantitative insight into the speciation of complex dynamic libraries of cages in the presence of small guest molecules was hence gained.

Overall, MS has been shown to be a valuable tool for the study of complex dynamic self-sorting mixtures. However, the general study of systems by mass spectrometric techniques is still limited due to a number of constraining factors such as the variability in the ionisation of each structure, the overlap of signals for structural isomers, and fluctuation in detection levels between instruments.

6.6 Experimental section

6.6.1 Synthesis of subcomponents 6.A and 6.B

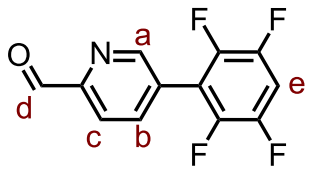
Synthesis of 5-phenylpicolinaldehyde 6.A

Phenylboronic acid (98.3 mg, 0.806 mmol, 1.50 eq.), 5-bromopicolinaldehyde (100 mg, 0.538 mmol, 1.00 eq.) and potassium phosphate (286 mg, 1.35 mmol, 2.50 eq.) were dissolved in 1,4-dioxane (5 mL) in a round bottom flask and degassed with nitrogen for 15 min. $Pd(PPh_3)_4$ (25 mg, 2.16 mmol, 0.040 eq.) was added and the mixture was degassed for a further 5 min. The round bottom flask was equipped with a condenser attached to a nitrogen line and the mixture was heated at 100 °C for 24 h, after which TLC analysis indicated the absence of starting material in the reaction mixture. The reaction was cooled to room temperature, diluted with water (50 mL) and extracted with CH_2Cl_2 (5 × 25 mL). The combined organic phases were dried over $MgSO_4$, filtered and the solvent removed *in vacuo* to give the crude product as a brown solid. Purification by silica gel chromatography (EtOAc in hexane, 10% - 100%) gave aldehyde **6.A** as a yellow solid (54.2mg, 0.296 mmol, 55% yield).

^{1}H NMR (400 MHz, 298K, $CDCl_3$) δ 10.14 (s, 1H), 9.02 (s, 1H), 8.10 – 8.01 (m, 2H), 7.65 (d, *J* = 7.1 Hz, 2H), 7.51 (m, 3H).

This data matches the data reported for the commercial product (CAS RN: 780800-85-3) from Acros Organics.

Synthesis of (2,3,5,6-tetrafluorophenyl)picolinaldehyde 6.B



(6-(1,3-Dioxolan-2-yl)pyridin-3-yl)boronic acid[22] (100 mg, 0.513 mmol, 1.00 eq.), 1-bromo-2,3,5,6-tetrafluorobenzene (176 mg, 0.770 mmol, 1.50 eq.) and potassium phosphate (272 mg, 1.28 mmol, 2.50 eq.) were dissolved in 1,4-dioxane/water (3:1 v/v, 6.7 mL) in a round bottom flask and degassed with nitrogen for 15 min. $Pd(PPh_3)_4$ (23.7 mg, 0.021 mmol, 0.040 eq.) was added and the mixture was degassed for a further 5 min. The round bottom flask was equipped with a condenser attached to a nitrogen line and the mixture was heated at 100 °C for 24 h, after which TLC analysis indicated the absence of starting material in the reaction mixture. The reaction was cooled to room temperature, diluted with water (50 mL) and extracted with CH_2Cl_2 (3 × 50 mL). The combined organics were dried over $MgSO_4$, filtered, and the solvent removed *in vacuo*. Purification through a plug of silica (EtOAc in hexane, 15%

- 100%) afforded a yellow solid, which was used directly in the next step. The solid was dissolved in THF/1M HCl(1:1) (30 mL). The reaction was heated at 60 °C for 18 h. The reaction was cooled to room temperature and the THF was removed *in vacuo*. The reaction was neutralised by the addition of 15 mL of an aqueous saturated solution of $NaHCO_3$ (15 mL). At this point the product precipitated out and was collected *via* centrifuge. The cream precipitate was washed with H_2O (2 × 10 mL) before being dried under a flow of nitrogen to furnish aldehyde **6.B** as a yellow solid (39.9 mg, 0.154 mmol, 30%).

^{1}H NMR (500 MHz, 298K, CD_3CN) δ 10.07 (s, 1H, H_d), 8.92 (d, *J* = 0.9 Hz, 1H, H_a), 8.13 (dd, *J* = 8.1, 0.9 Hz, 1H, H_b), 8.08 (dd, *J* = 8.1, 0.9 Hz, 1H, H_c), 7.47 (tt, *J* = 10.2, 7.6 Hz, 1H, H_e). **^{19}F NMR** (376 MHz, 298K, CD_3CN) δ−140.5 (dd, *J* = 21.5, 12.8 Hz), −145.0 (dd, *J* = 21.4, 12.8 Hz). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 192.7 (s), 152.6 (s), 150.9 (s), 147.8 − 147.4 (m), 145.3 − 144.8 (m), 142.6 – 142.3 (m), 138.7 (s), 128.2 (s), 121.4 (s), 117.0 (s), 106.8 (s). **TOF-MS-ES+**: Calculated [**B**+H]$^+$=256.0385 Observed [**B**+H]$^+$=256.0363

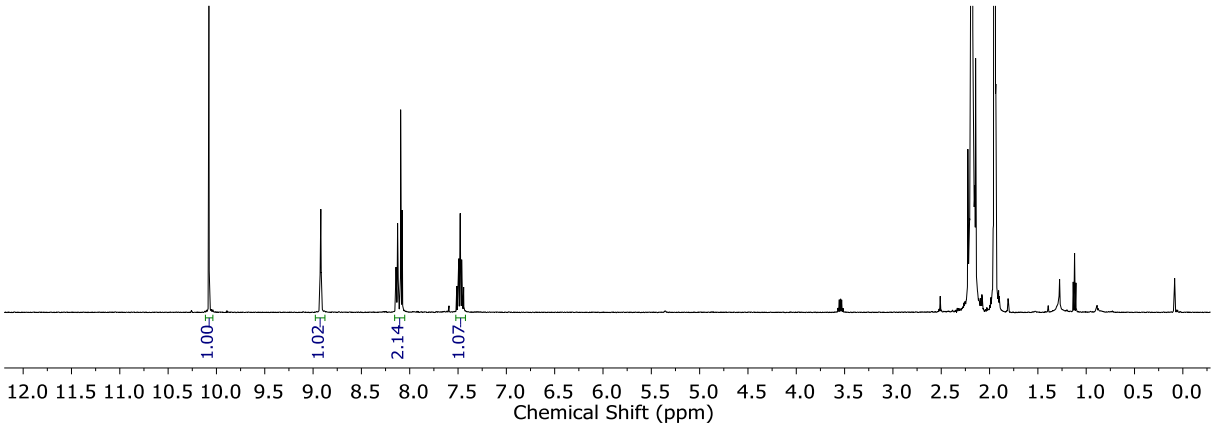


Figure 6.34 | ^{1}H NMR (400 MHz, 298K, CD_3CN) of **6.B**

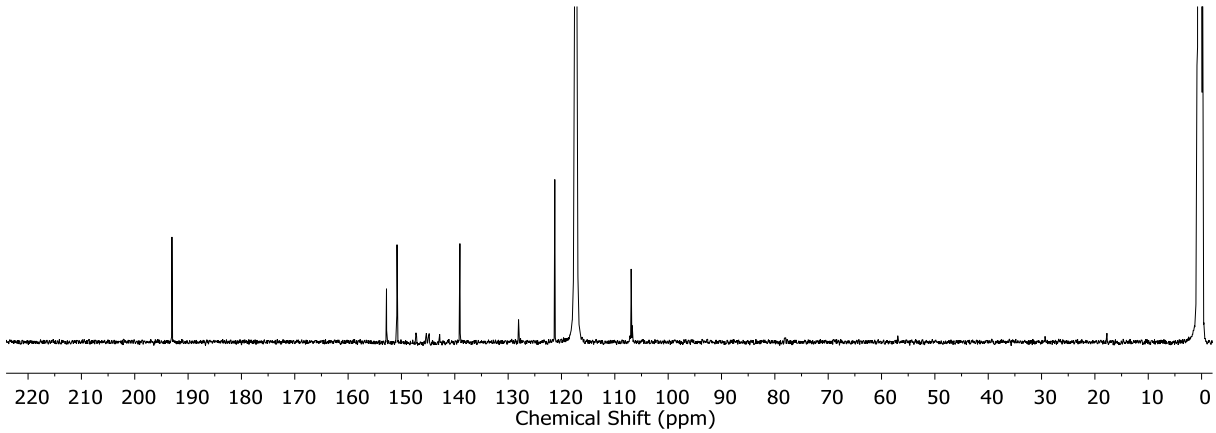


Figure 6.35 | ^{13}C NMR (126 MHz, 298K, CD_3CN) of **6.B**.

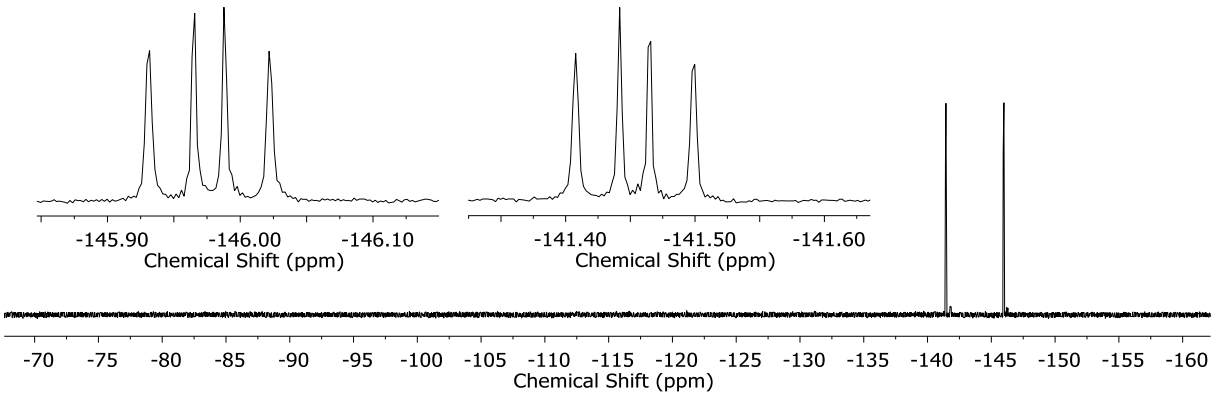
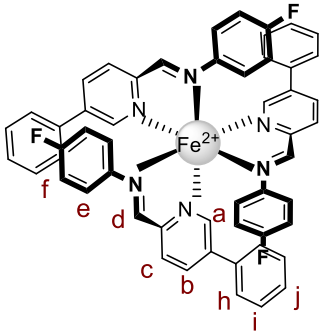


Figure 6.36 | ^{19}F NMR (376 MHz, 298K, CD_3CN) of **6.B**

6.6.2 Synthesis of mononuclear complexes 6.1 and 6.2

Mononuclear complex 6.1



5-Phenylpicolinaldehyde **6.A** (2.0 mg, 11.0 µmol, 3.0 eq.), *p*-fluoroaniline (1.2 mg, 11.0 µmol, 3.0 eq.) and $Fe(NTf_2)_2.6H_2O$ (2.5 mg, 3.60 µmol, 1.0 eq.) were dissolved in CH_3CN (2 mL) in a sealed 5 mL reaction tube. The solution was heated at 50 °C for 24 h. The dark purple solution was then cooled and concentrated under a flow of nitrogen. The solution was dried under a flow of nitrogen and redissolved in CD_3CN for analysis.

Mononuclear complex **6.1** was obtained as a mixture of *fac* and *mer* isomers.

^{1}H NMR (400 MHz, 298K, CD_3CN) δ, *mer*-**1** isomer: 9.53 (s, 1H, *mer*-H_d), 9.41 (s, 1H, *mer*-H_d), 9.16 (s, 1H, *mer*-H_d), 8.75 (s, 1H, *mer*-H_a), 8.65 (s, 1H, *mer*-H_c), 8.63 (s, 1H, *mer*-H_c), 8.61 (s, 1H, *mer*-H_a), 8.44 (s, 1H, *mer*-H_b), 8.25 (s, 1H, *mer*-H_b), 8.14 (s, 1H*mer*-H_c), 8.09 (s, 1H, *mer*-H_b), 8.00 (s, 1H, *mer*-H_a), 7.63 – 7.45 (m, 30H, 3 *mer*-H_h + 3 *mer*-H_i + 3 *mer*-H_j + *fac*-H_h + *fac*-H_i + *fac*-H_j), 7.22 (s, 2H, *mer*-H_f), 6.95 (s, 2H, *mer*-H_f), 6.88 (s, 2H, *mer*-H_e), 6.78 (s, 4H, *mer*-H_f+*mer*-H_e), 6.15 (s, 2H, *mer*-H_e). **^{1}H NMR** (400 MHz, 298K, CD_3CN) δ, *fac*-**1** isomer: 9.13 (s, 3H, *fac*-H_d), 8.60 (s, 3H, *fac*-H_b),8.12 (s, 3H, *fac*-H_c), 7.78 (s, 3H, *fac*-H_a),7.63 – 7.45 (m, 30H, 3 *mer*-H_h + 3 *mer*-H_i +3 *mer*-H_j + *fac*-H_h + *fac*-H_i +*fac*-H_j),7.07 (t, J = 8.5 Hz, 6H, *fac*-H_f),5.65 (dd, J = 8.4, 4.6 Hz, 6H, *fac*-H_e). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 175.7 (s), 174.9 (s), 173.6 (s), 172.7(s), 157.5 – 155.0 (m), 143.5 (s), 138.5 – 137.7 (m), 136.8 – 135.3 (m), 133.0 – 131.6 (m), 131.2 – 130.1 (m), 129.1 – 128.1 (m), 126.0 – 124.5 (m), 122.3 – 122.0 (m), 119.9 – 119.0 (m), 117.6 – 115.7 (m). **^{19}F NMR** (376 MHz, 298K, CD_3CN) δ −80.60 (s,

NTf_2^-), −113.47 (s, *mer*-F + *mer*-F), −113.91 (s, *mer*-F), −114.20 (s, *fac*-F). **LR-ESI-MS**: [charge, calculated mass]: *m/z* = 442.3 [**6.1**$^{2+}$, 442.3], 1164.2 [**6.1**$(NTf_2)^+$, 1164.6]. **HR-ESI-MS**: *m/z* calculated for **6.1**$^{2+}$ = 442.1263, observed = 442.1260.

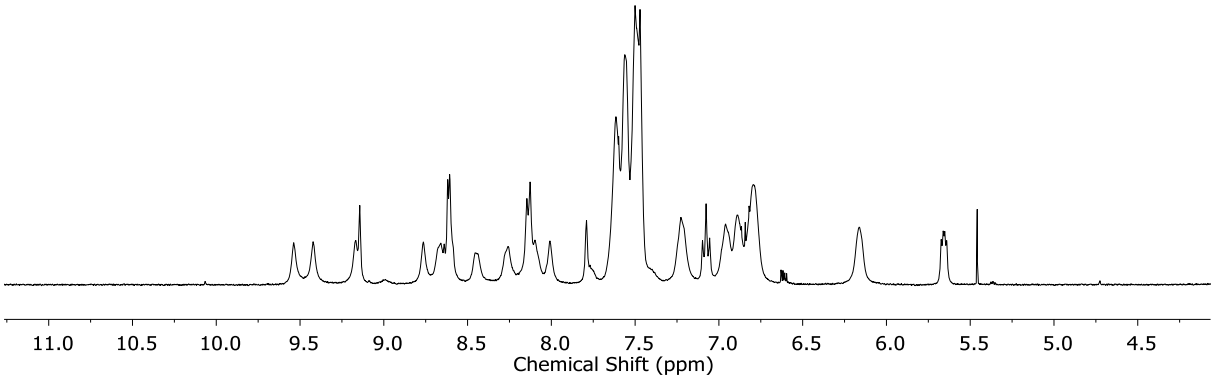


Figure 6.37 | Aromatic region of the ^{1}H NMR spectrum (400 MHz, 298K, CD_3CN) of a solution of complex **6.1**

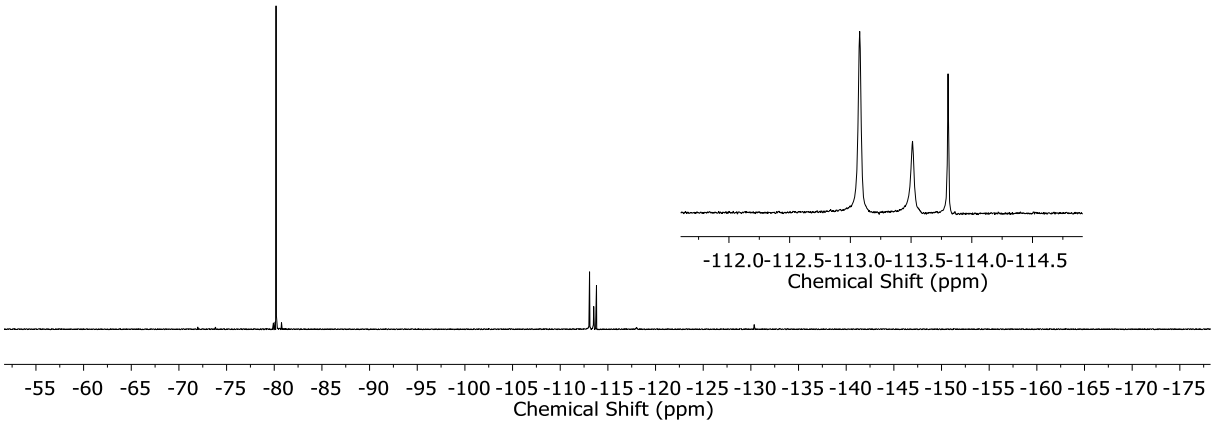


Figure 6.38 | ^{19}FNMR spectrum (376 MHz, 298K, CD_3CN) of a solution of complex **6.1**

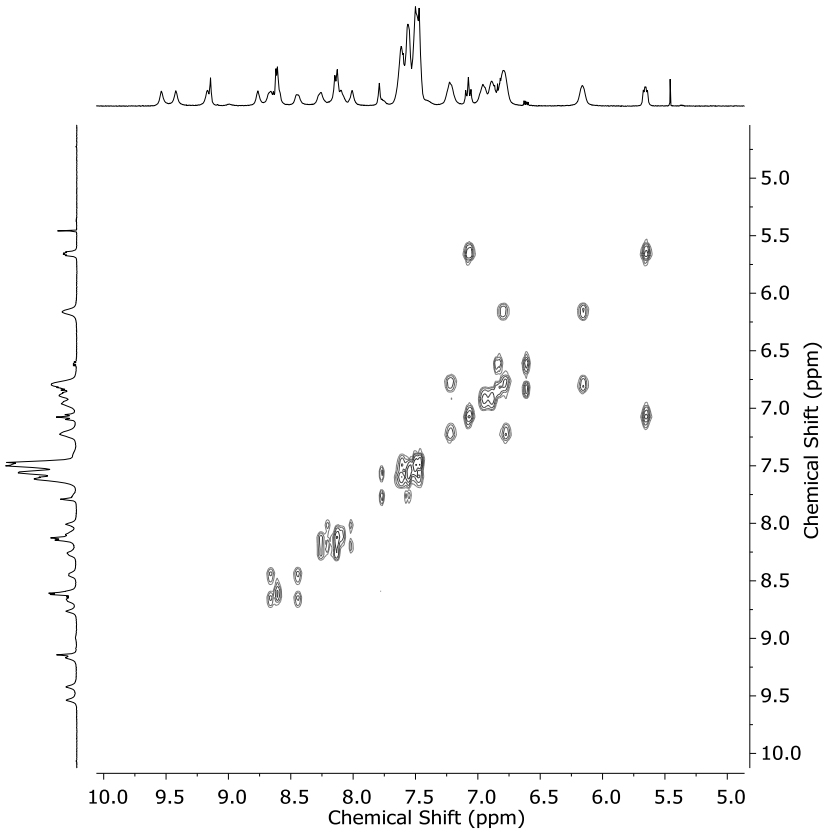


Figure 6.39 | Aromatic region of the ^{1}H-^{1}H COSY spectrum (500 MHz, 298K, CD_3CN) of a solution of complex **6.1.**

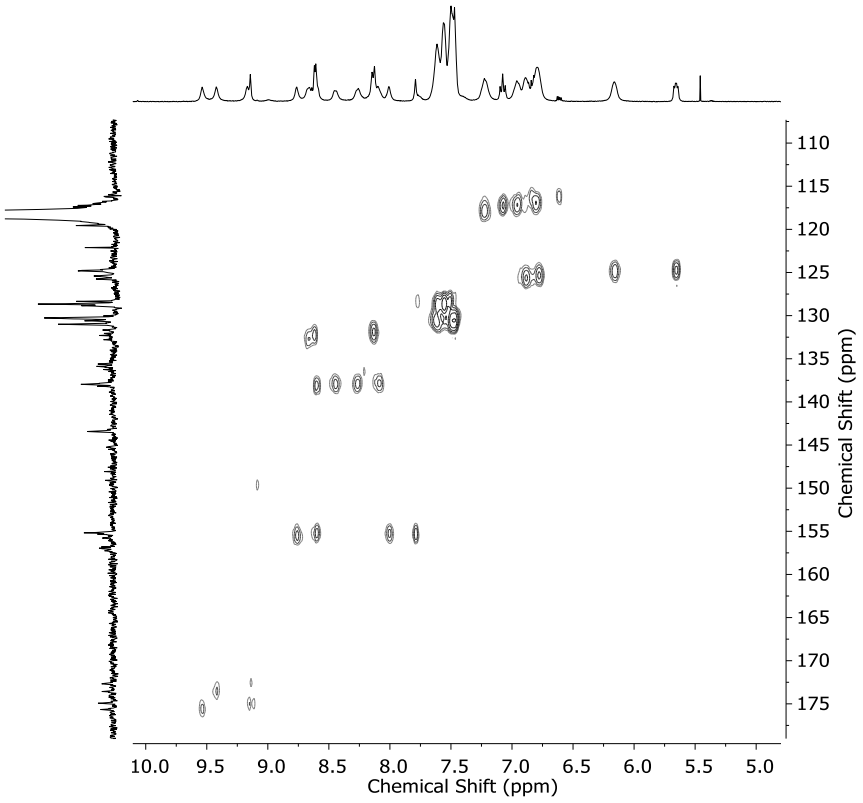
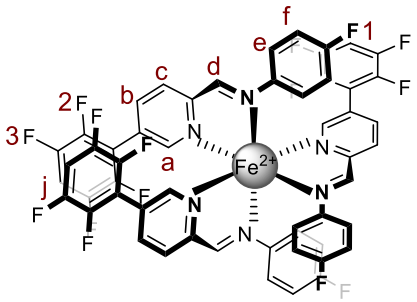


Figure 6.40 | Aromatic region of the ^{1}H-^{13}C HSQC spectrum (500 MHz, 298K, CD_3CN) of a solution of complex **6.1**

Mononuclear complex 6.2



(2,3,5,6-Tetrafluorophenyl)picolinaldehyde **6.B** (2.0 mg, 7.80 µmol, 3.0 eq.), *p*-fluoroaniline (0.87 mg, 7.80 µmol, 3.0 eq.) and $Fe(NTf_2)_2.6H_2O$ (1.8 mg, 2.60 µmol, 1.0 eq.) were dissolved in CH_3CN (2 mL) in a sealed 5 mL reaction tube. The solution was heated at 50 °C for 24 h. The dark purple solution was then cooled and concentrated under a flow of nitrogen.

^{1}H NMR (500 MHz, 298K, CD_3CN) δ 9.42 (s, 1H, H_d), 9.20 (s, 1H, H_d), 9.04 (s, 1H, H_d), 8.67 (d, *J* = 8.1 Hz, 1H, H_c), 8.65 (s, 1H, H_a), 8.45 (s, 1H, H_a), 8.43 (d, *J* = 7.5 Hz, 1H, H_b), 8.26 (d, *J* = 7.5 Hz, 1H, H_b), 8.19 (d, *J* = 8.1 Hz, 1H, H_c), 8.16 (d, *J* = 8.1 Hz, 1H, H_c), 8.12 (d, *J* = 7.5 Hz, 1H, H_b), 7.88 (s, 1H, H_a), 7.63 – 7.39 (m, 3H, H_j), 7.19 (t, *J* = 7.5 Hz, 2H, H_f), 7.03 (t, 2H,*J* = 7.5 Hz, H_f), 6.86 (underneath *p*-fluoroaniline peak, 4H, H_e + H_f), 6.77 (b, 2H, H_e), 6.29 (b, 2H, H_e). **^{13}C NMR** (126 MHz, 298K, CD_3CN) δ 175.3 (s), 173.5 (s), 172.3 (s), 158.6 – 157.1 (m), 156.2 (s), 147.6 – 146.8 (m), 145.9 – 144.5 (m), 143.8 – 142.0 (m), 141.2 – 139.8 (m), 132.2 – 129.0 (m), 124.9 – 123.3 (m), 121.2 (s), 119.7 – 118.3 (m), 116.7 – 115.8 (m), 109.2 – 107.6 (m). **LR-ESI-MS**: [charge, calculated mass]: *m/z* = 549.9 [**6.2**$^{2+}$, 550.3], 1379.8 [**6.2**$(NTf_2)^+$, 1380.5] **HR-ESI-MS**: *m/z* calculated for **6.2**$^{2+}$ = 550.0698, observed = 550.0687.

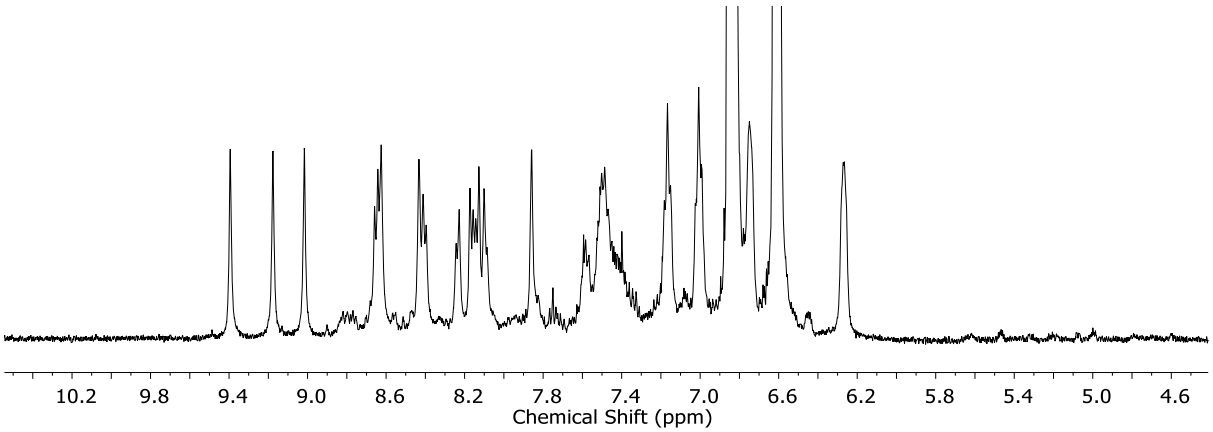


Figure 6.41 | Aromatic region of the^{1}H NMR spectrum (500 MHz, 298K, CD_3CN) of a solution of complex **6.2**

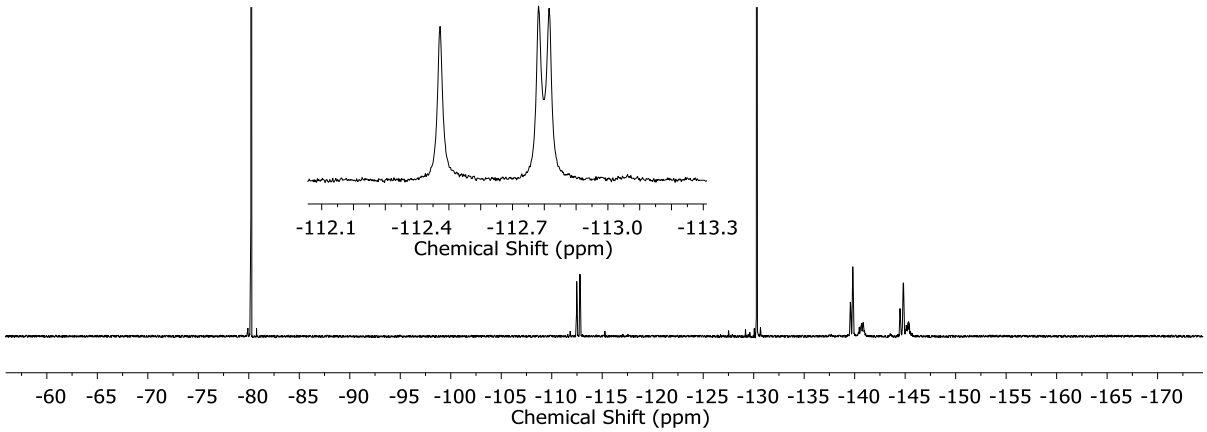


Figure 6.42 | ^{19}FNMR spectrum (376 MHz, 298K, CD_3CN) of a solution of complex **6.2**

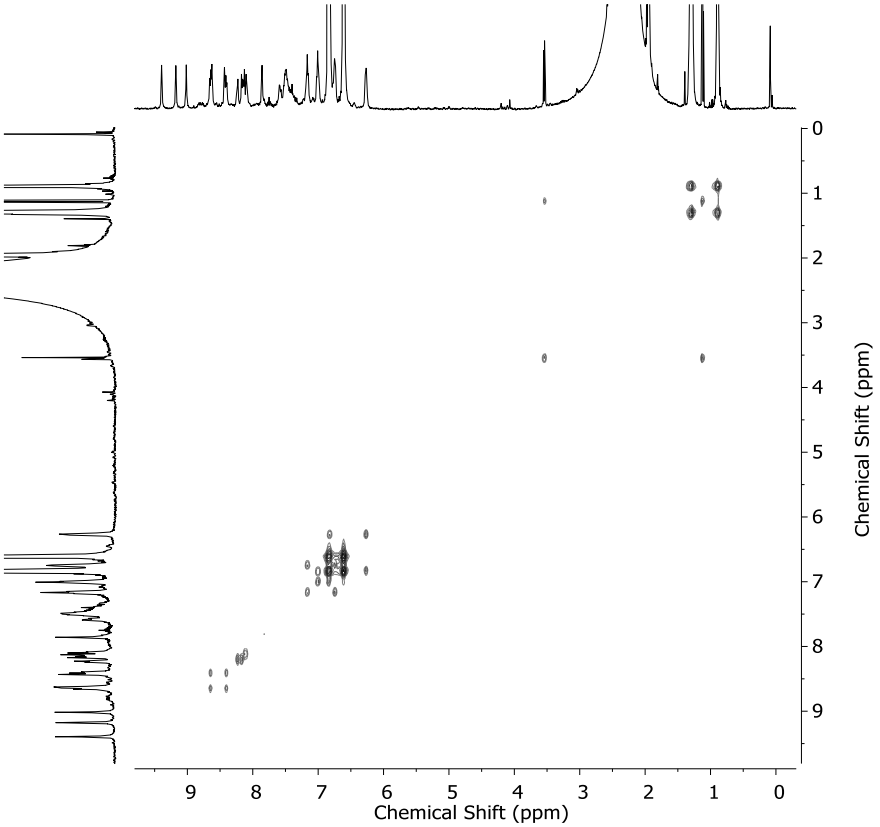


Figure 6.43 | Aromatic region of the ^{1}H-^{1}H COSY spectrum (500 MHz, 298K, CD_3CN) of a solution of complex **6.2**

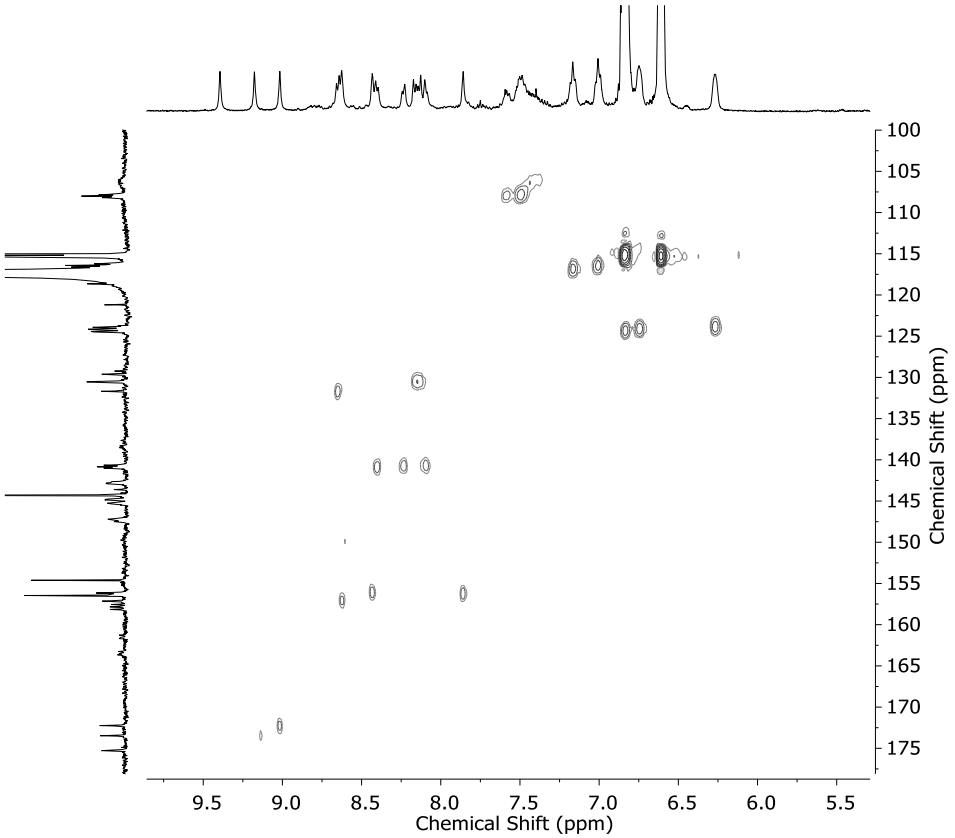
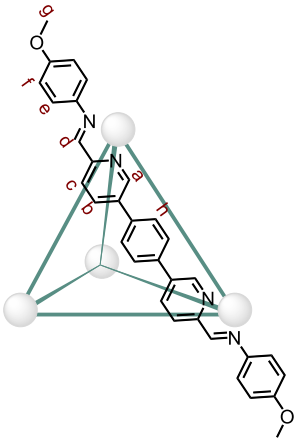


Figure 6.44 | Aromatic region of the ^{1}H-^{13}C HSQC spectrum (500 MHz, 298K, CD_3CN) of a solution of complex **6.2.**

6.6.3 Synthesis of cages 6.3 and 6.4

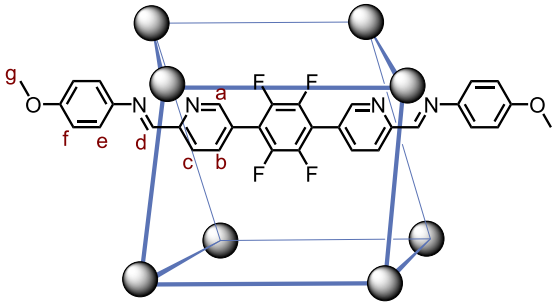
5,5'-(1,4-phenylene)dipicolinaldehyde **6.C** or 5,5'-(Perfluoro-1,4-phenylene)dipicolinaldehyde **6.D** (5.50 µmol, 3.0 eq.), *p*-methoxyaniline (1.4 mg, 11.0 µmol, 6.0 eq.) and $Fe(PF_6)_2.6CH_3CN$ (2.2 mg, 3.70 µmol, 2.0 eq.) were dissolved in CH_3CN (2 mL) in a sealed 5 mL reaction tube. The solution was heated at 50 °C for 24 h. The dark green or blue respectively solutions were then cooled and concentrated under a flow of nitrogen. Addition of 10 mL of diethyl ether precipitated the cages as a dark green or blue solid. The solid was separated by centrifugation and washed with diethyl ether (2×10 mL). The solid was then dried under a flow of nitrogen and dissolved in CD_3CN.

Cage 6.3



^{1}H NMR (400 MHz, 298K, CD_3CN) δ 8.83 (s, 12H, H_d), 8.67 (d, *J* = 8.1 Hz, 12H, H_b), 8.56 (d, *J* = 8.2 Hz, 12H, H_c), 7.50 (s, 24H, H_h), 7.45 (s, 12H, H_a), 6.78 (d, *J* = 8.8 Hz, 24H, H_e), 5.71 (d, *J* = 8.9 Hz, 24H, H_f), 3.81 (s, 36H, H_g). These data match those previously reported.[34]

Cage 6.4



^{1}H NMR (500 MHz, 298K, CD_3CN) δ: 9.14 (s, 8H, H_d), 8.92 (s, 8H, H_d), 8.87 (s, 8H, H_d), 8.66 (d, *J* = 8.6 Hz, 8H, H_c), 8.35 (s, 8H, H_a), 8.32 (d, *J* = 8.7 Hz, 8H, H_b), 8.28 – 8.20 (m, 16H, 2 × H_b), 8.13 (d, *J* = 8.0 Hz, 8H, H_c),8.05 (d, *J* = 8.0 Hz, 8H, H_b), 7.97 (d, *J* = 8.1 Hz, 8H, H_c), 7.82 (s, 8H, H_a), 6.86 (d, *J* = 9.2 Hz, 16H, H_f), 6.76 (d, *J* = 9.0 Hz, 16H, H_e), 6.73 – 6.65 (m, 32H, H_e + H_f), 6.53 (d, *J* = 8.8 Hz, 16H, H_f), 6.25 (d, *J* = 8.8 Hz, 16H, H_e), 3.73 (s, 24H, H_g), 3.71 (s, 24H, H_g),3.65 (s, 24H, H_g). **^{19}F NMR** (376 MHz, 298K, CD_3CN) δ: −72.72 (d, *J* = 708.6 Hz, PF_6^-), −143.70 (m), −144.23 (s), −144.59 – −145.05 (m). These data match those previously reported.[22]

6.6.4 Sorting experiments for cages 6.3 and 6.4

Preparation of a mixture of 6.3 and 6.4 from subcomponents.

5,5'-(1,4-Phenylene)dipicolinaldehyde **6.C** (0.80 mg, 2.80 µmol, 3.0 eq.), 5,5'-(perfluoro-1,4-phenylene)dipicolinaldehyde **6.D** (1.0 mg, 2.80 µmol, 3.0 eq.), *p*-anisidine (1.4 mg, 11.0 µmol, 12 eq.) and $Fe(PF_6)_2.6CH_3CN$ (2.2 mg, 3.70 µmol, 4.0 eq.) were dissolved in CD_3CN (0.7 mL) in a sealed J.Young NMR tube. The reaction mixture was kept at 50 °C for 3 days, yielding a dark blue/green solution.

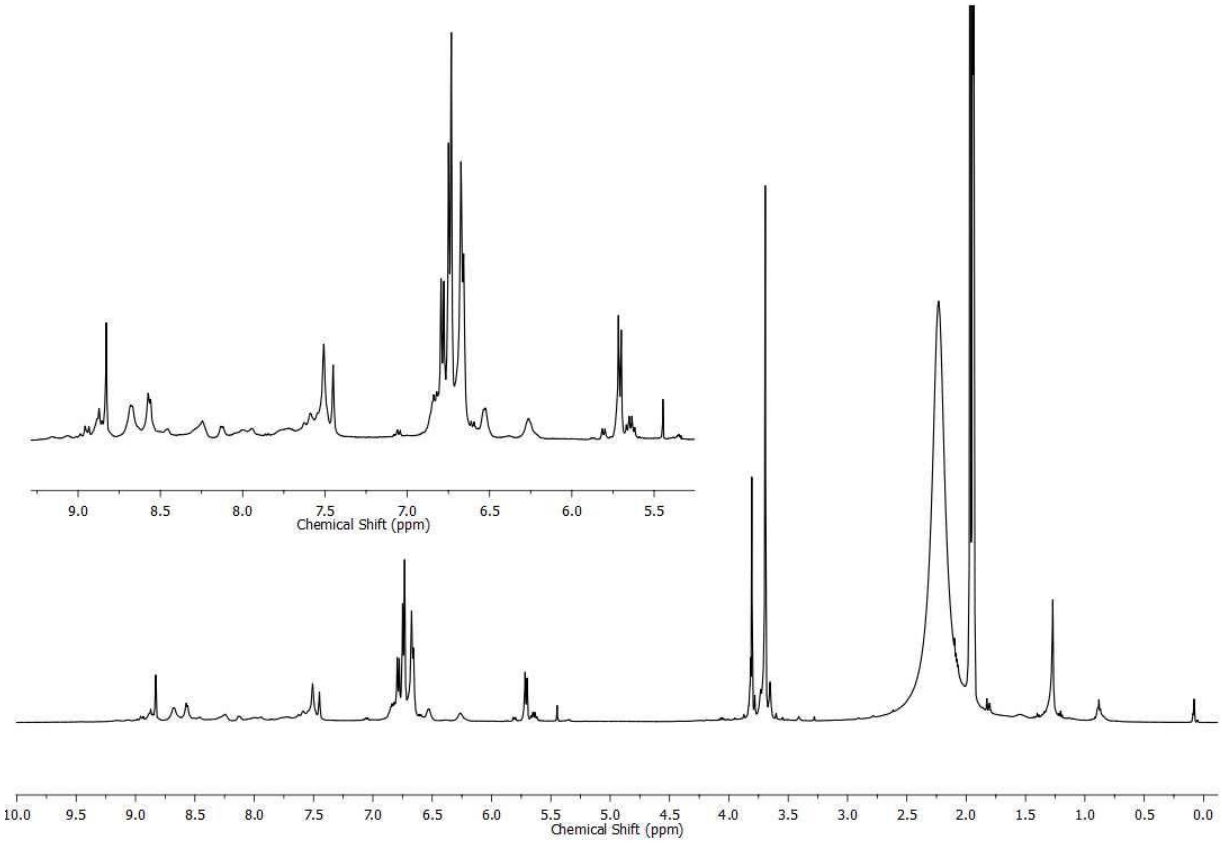


Figure 6.45 | ^{1}H NMR (500 MHz, 298K, CD_3CN) spectrum of a mixture of **6.3** and **6.4** prepared from subcomponents after 3 days at 50 °C.

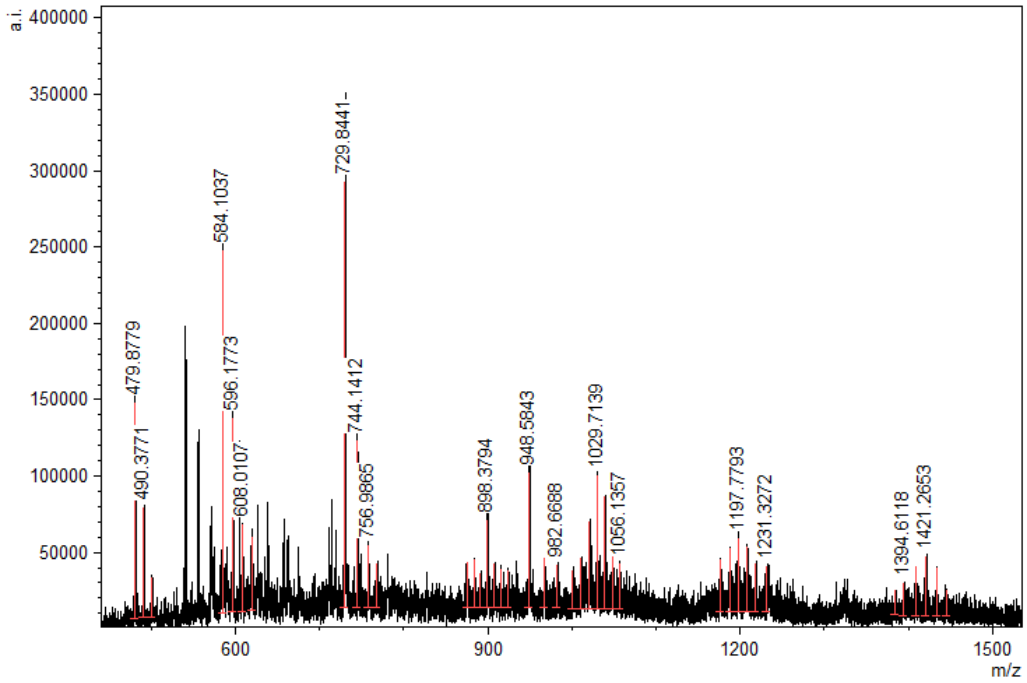


Figure 6.46 | ESI-Mass spectrum of a mixture of **6.3** and **6.4** prepared from subcomponents after 3 days at 50 °C.

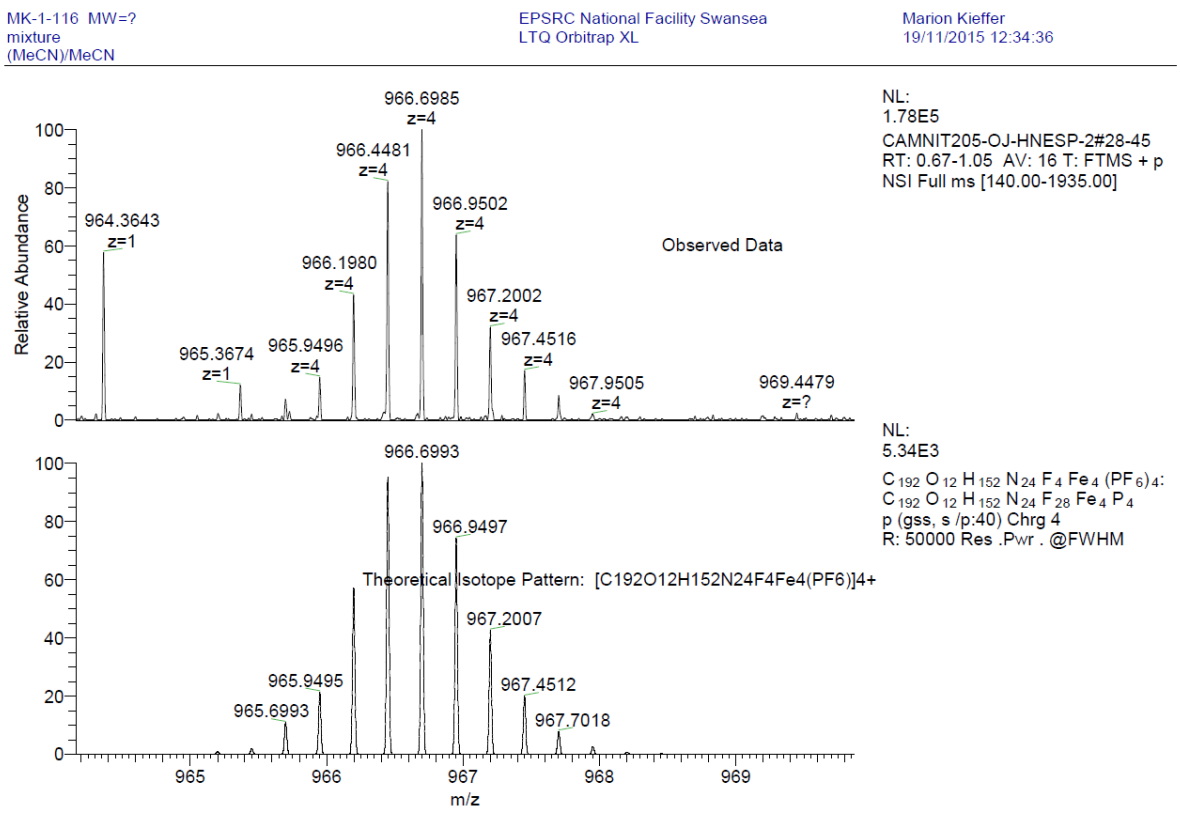


Figure 6.47 | HR-ESI-Mass spectrum and theoretical isotope pattern of [**6.3**$(PF_6)_4$]$^{+4}$ after exchange of one ligand with the prismatic structure.

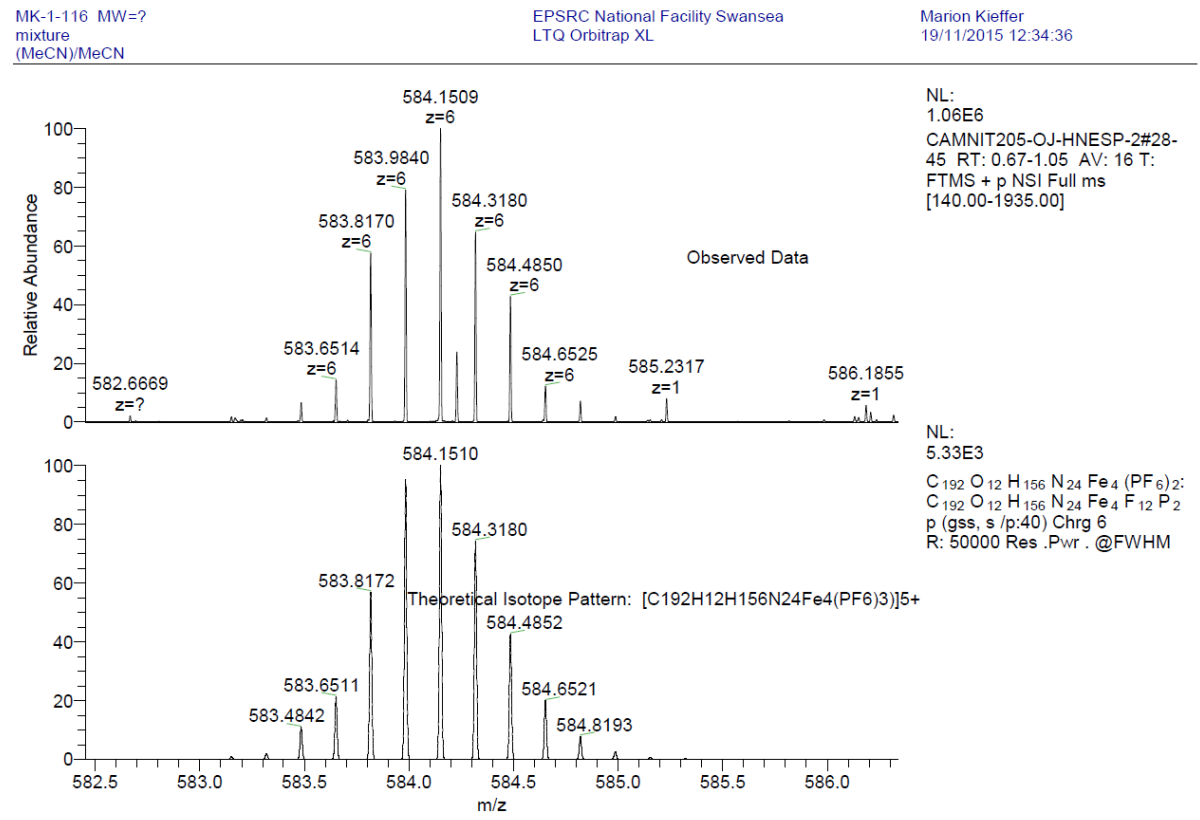


Figure 6.48 | HR-ESI-Mass spectrum and theoretical isotope pattern of [**6.3**$(PF_6)_2$]$^{+6}$.

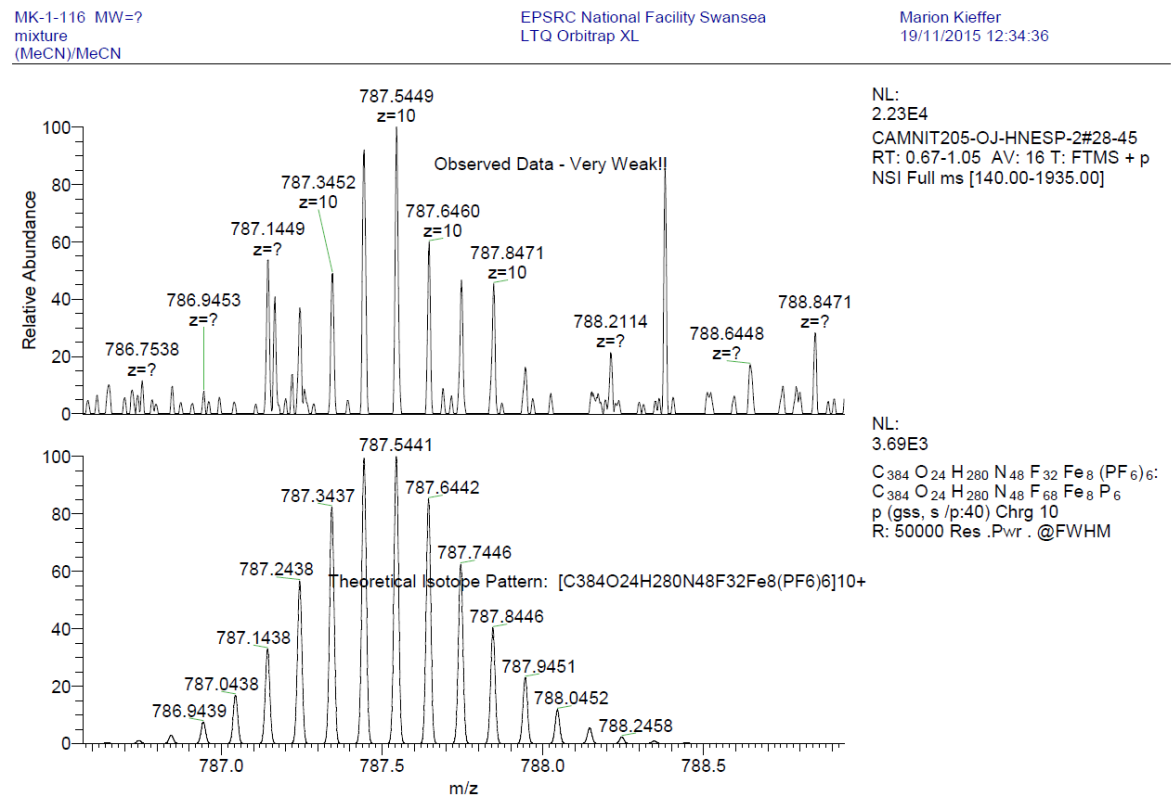


Figure 6.49 | HR-ESI-Mass and theoretical isotope pattern of $[\mathbf{6.4}(PF_6)_6]^{+10}$ after exchange of four ligands with the tetrahedron.

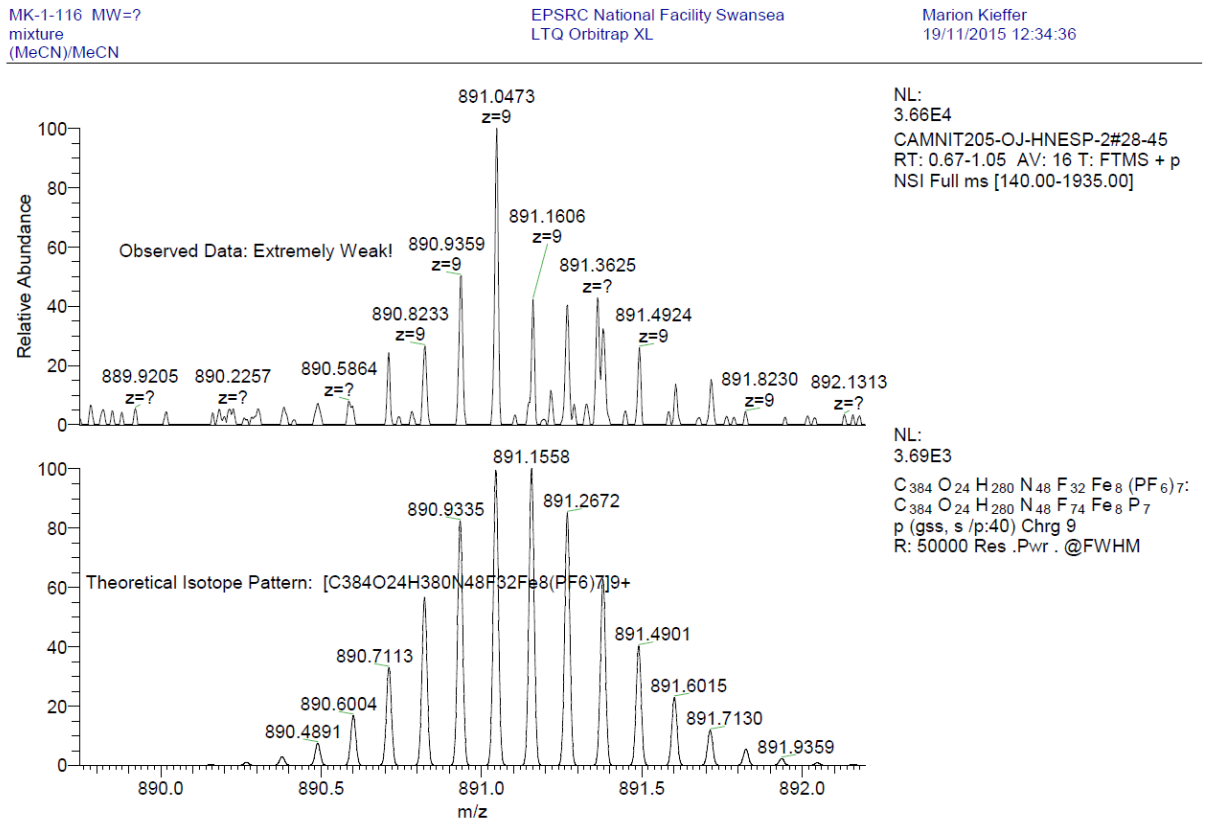


Figure 6.50 | HR-ESI-Mass and theoretical isotope pattern of $[\mathbf{6.4}(PF_6)_7]^{+9}$ after exchange of four ligands with the tetrahedron.

Preparation of a mixture of 6.3 and 6.4 from pre-assembled cages.

6.3 (3.5 mg, 0.82 µmol, 2.0 eq.) and **6.4** (3.9 mg, 0.41 µmol, 1.0 eq.) were dissolved in CD_3CN (0.7 mL) in a sealed J. Young NMR tube. The reaction mixture was stirred at 50 °C for 1 week, yielding a dark blue/green solution.

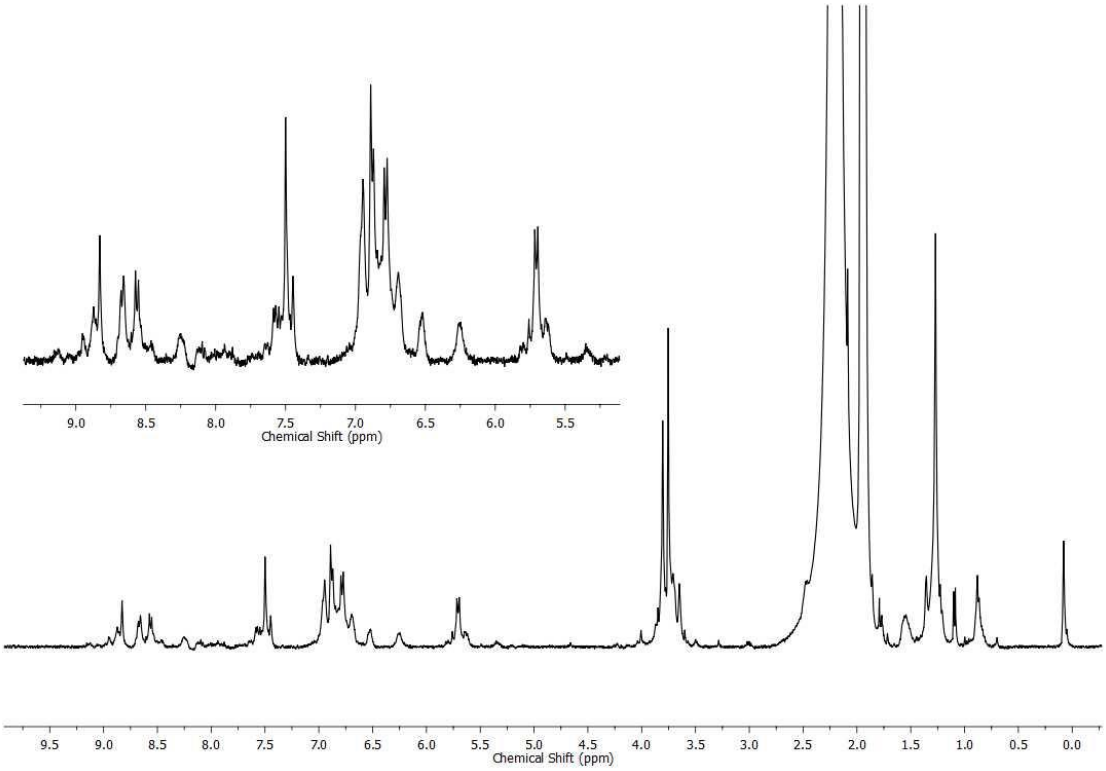


Figure 6.51 | ^{1}H NMR (400 MHz, 298K, CD_3CN) spectrum of a mixture of **6.3** and **6.4** prepared from pre-assembled cages after 3 days at 50 °C.

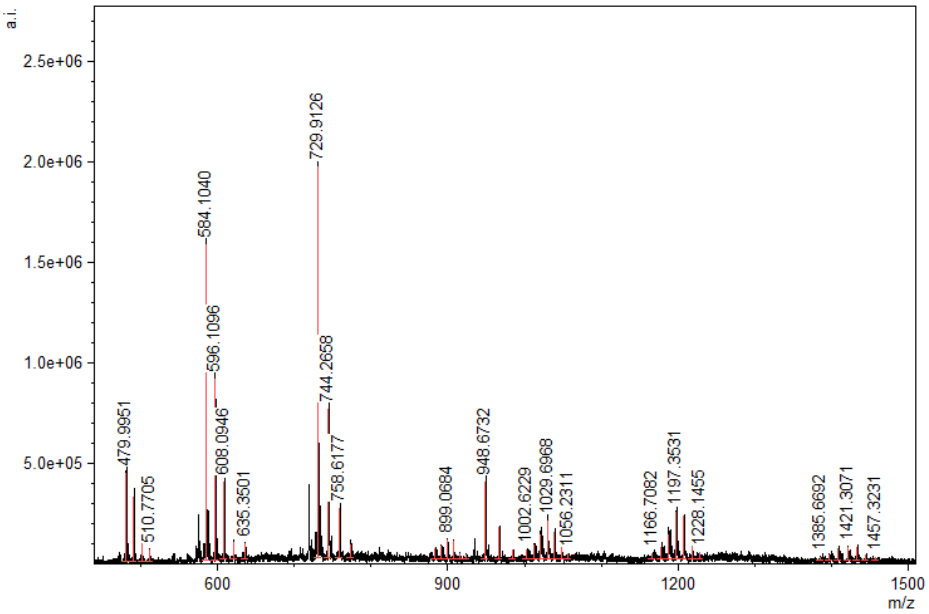


Figure 6.52 | ESI-Mass spectrum of a mixture of **6.3** and **6.4** prepared from pre-assembled cages after 3 days at 50 °C.

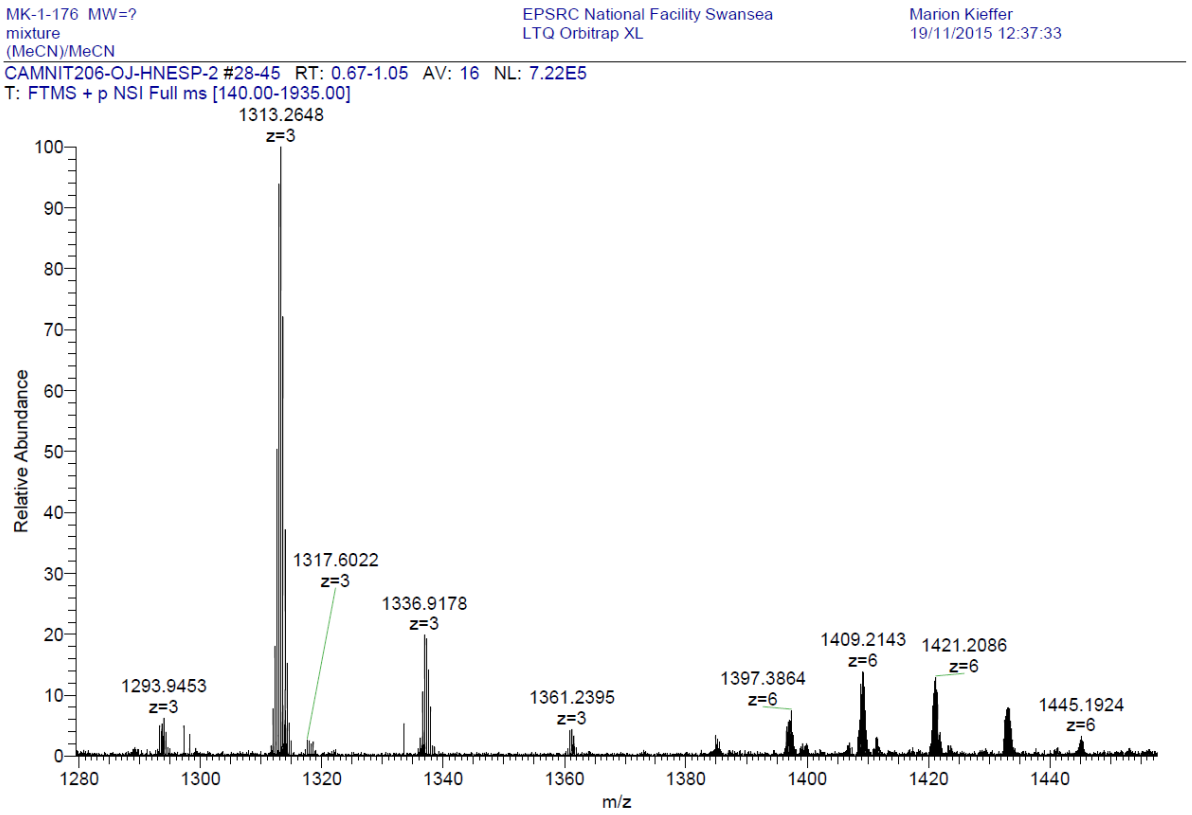


Figure 6.53 | HR-ESI-Mass spectrum and showing the peaks of $[\mathbf{6.3}(PF_6)_5]^{+3}$ and the structures formed by subsequent exchange of ligand with **6.4** and the peaks of $[\mathbf{6.4}(PF_6)_{10}]^{+6}$ and the structures formed by subsequent exchange of ligand with **6.3**.

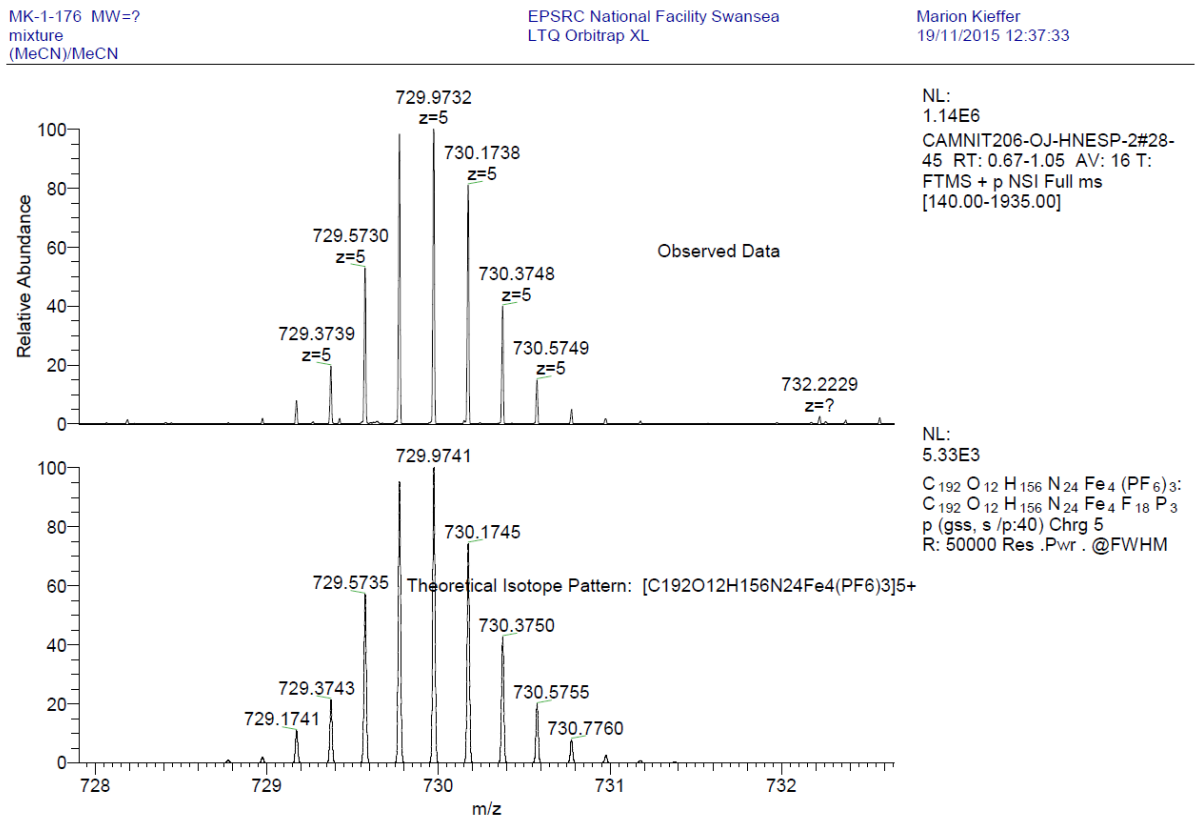


Figure 6.54 | HR-ESI-Mass spectrum and theoretical isotope pattern of $[\mathbf{6.3}(PF_6)_3]^{+5}$

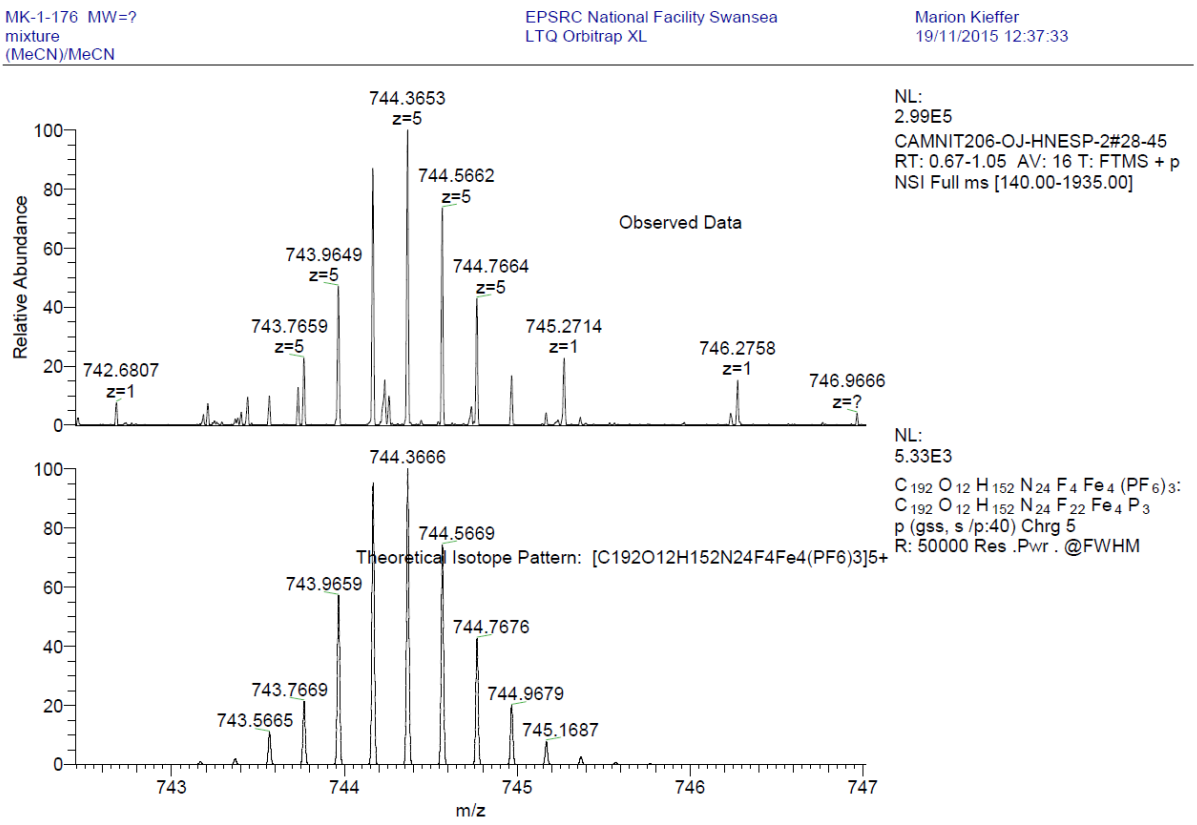


Figure 6.55 | HR-ESI-Mass spectrum and theoretical isotope pattern of $[\mathbf{6.3}(PF_6)_3]^{+5}$ after exchange of one ligand with the prismatic structure.

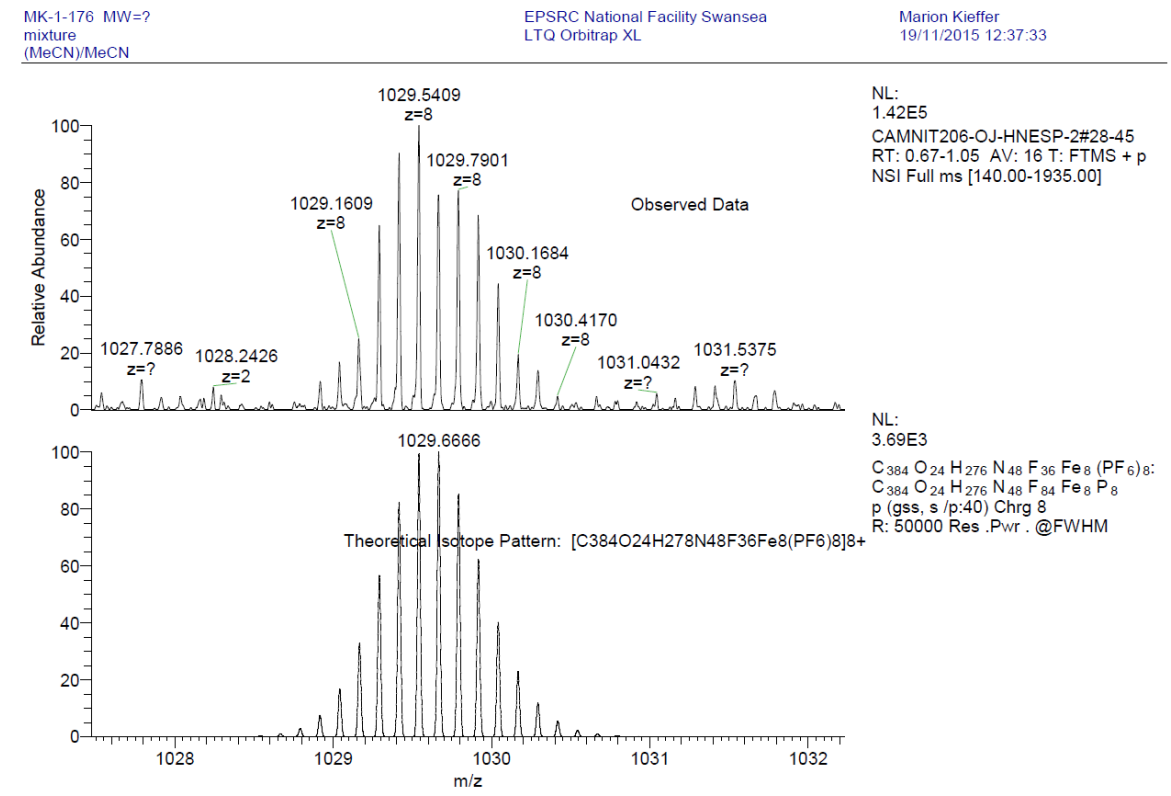


Figure 6.56 | HR-ESI-MS spectrum and theoretical isotope pattern of $[\mathbf{6.4}(PF_6)_8]^{+8}$ after exchange of three ligands with the tetrahedron.

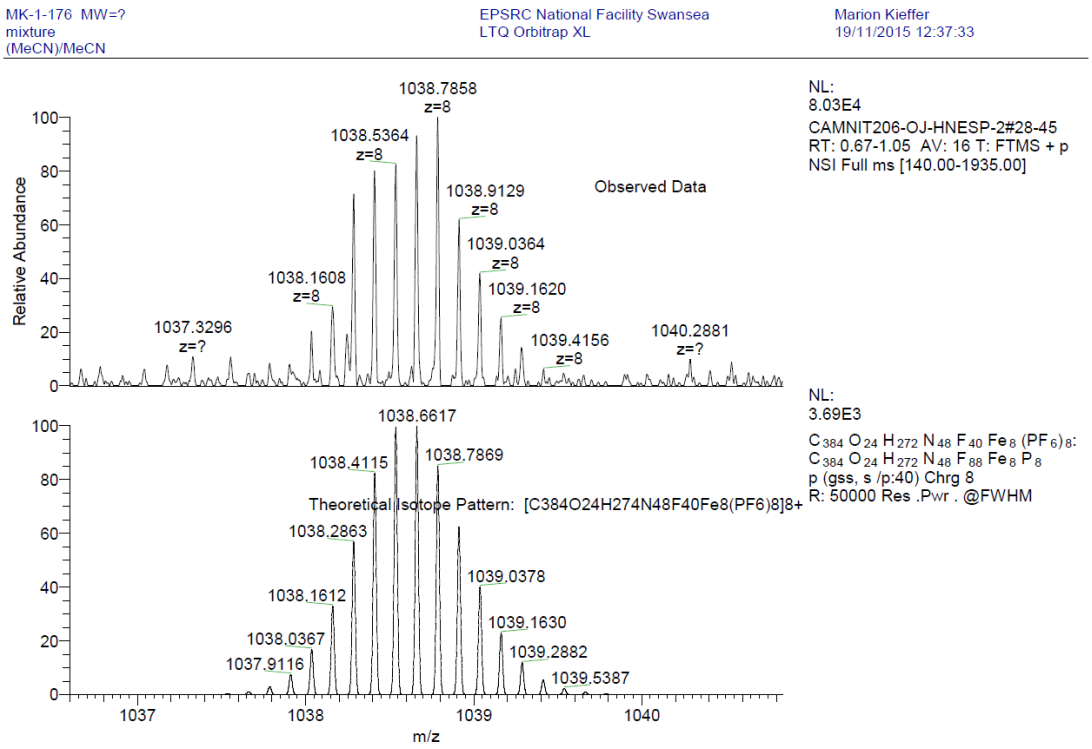


Figure 6.57 | HR-ESI-Mass spectrum and theoretical isotope pattern of [**6.4**$(PF_6)_8$]$^{+8}$ after exchange of two ligands with the tetrahedron.

6.6.5 Synthesis and characterization of cages 6.5 – 6.8

Cages **6.5 – 6.8** were synthesised and characterised by Felix Rizzuto. All data, including crystallographic data, can be found in the literature.[23]

6.6.6 Sorting experiments for cages 6.5 – 6.8

General procedures

- Two triamines (1.60 µmol of each triamine, 4 eq. of each triamine), 2-formylphenanthroline **6.G** (2.00 mg, 9.61 µmol, 24 eq.) and $Co(OTf)_2$ (1.71 mg, 4.80 µmol, 12 eq.) were combined in CD_3CN (0.5 mL) and heated at 70 °C for 18 h after which equilibrium was reached and the self-sorted output was formed.
- Stock solutions of **6.5**, **6.6**, **6.7** and **6.8** (0.89 mM) were respectively prepared from triamine **6.H**, **6.I**, **6.J** and **6.K** (3.20 µmol, 4 eq.), 2-formylphenanthroline (2.00 mg, 9.61 µmol, 12 eq.) and $Co(OTf)_2$ (1.72 mg, 4.81 µmol, 6 eq.) and heated at 70 °C for 18 h in CH_3CN (0.9 mL). Once cooled to room temperature, 0.1 mL of two homoleptic cage solutions were combined, making solutions of two purely homoleptic complexes: **6.5** and **6.6**, **6.5** and **6.7**,

6.5 and **6.8**, **6.6** and **6.7**, **6.6** and **6.8** and **6.7** and **6.8**. These mixtures were heated at 70 °C for 18 h, after which equilibrium was reached and the self-sorted output was formed.

Note: NMR data of the sorting mixtures was acquired by Felix Rizzuto. IM-MS data was acquired by Felix Rizzuto and Marion Kieffer. ESI-MS data was acquired by Marion Kieffer.

Subcomponents 6.H and 6.I

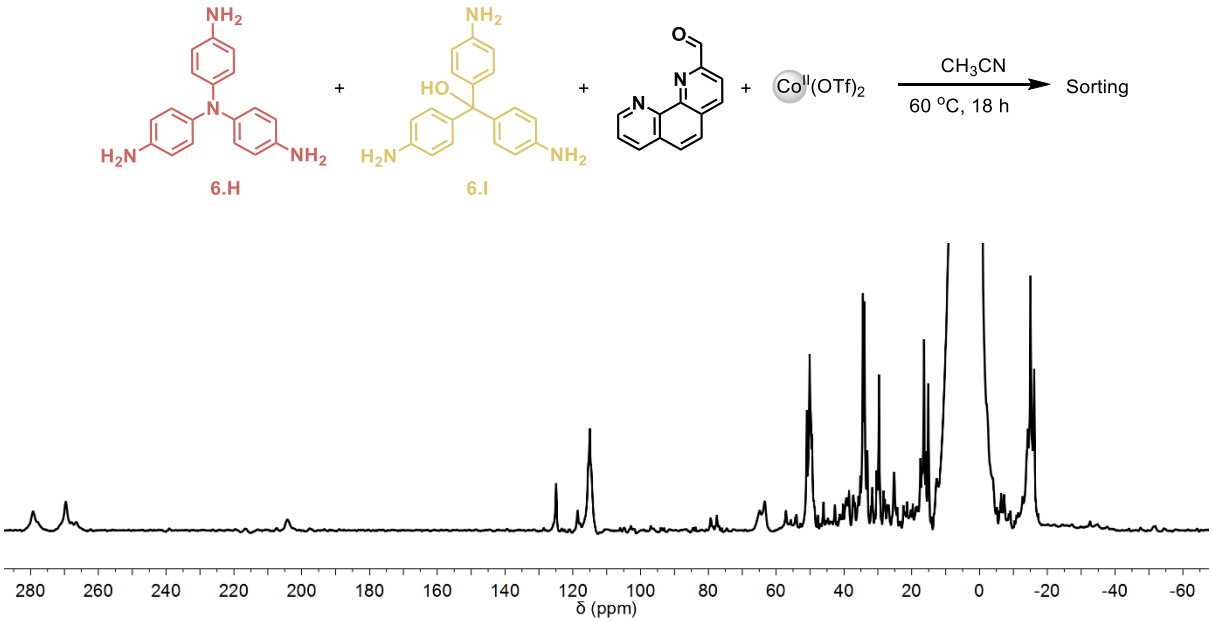


Figure 6.58 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of the mixture of cages generated when both **6.H** and **6.I** were used in the self-assembly process.

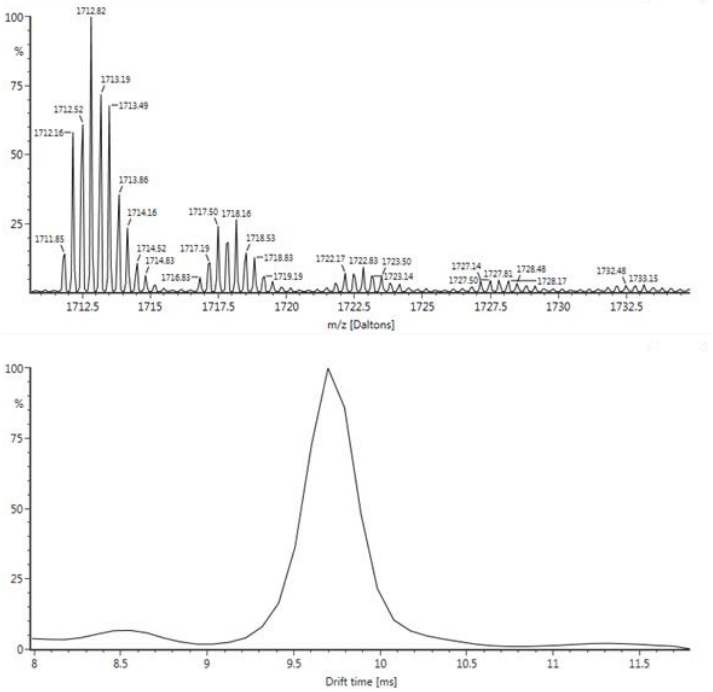


Figure 6.59 | IM-MS data for cages generated when both **6.H** and **6.I** were used in the self-assembly process, showing the z = +3 peaks (top). The assembly resolved a single, broad drift time, indicating similar sizes between the sorted species.

Subcomponents 6.H and 6.J

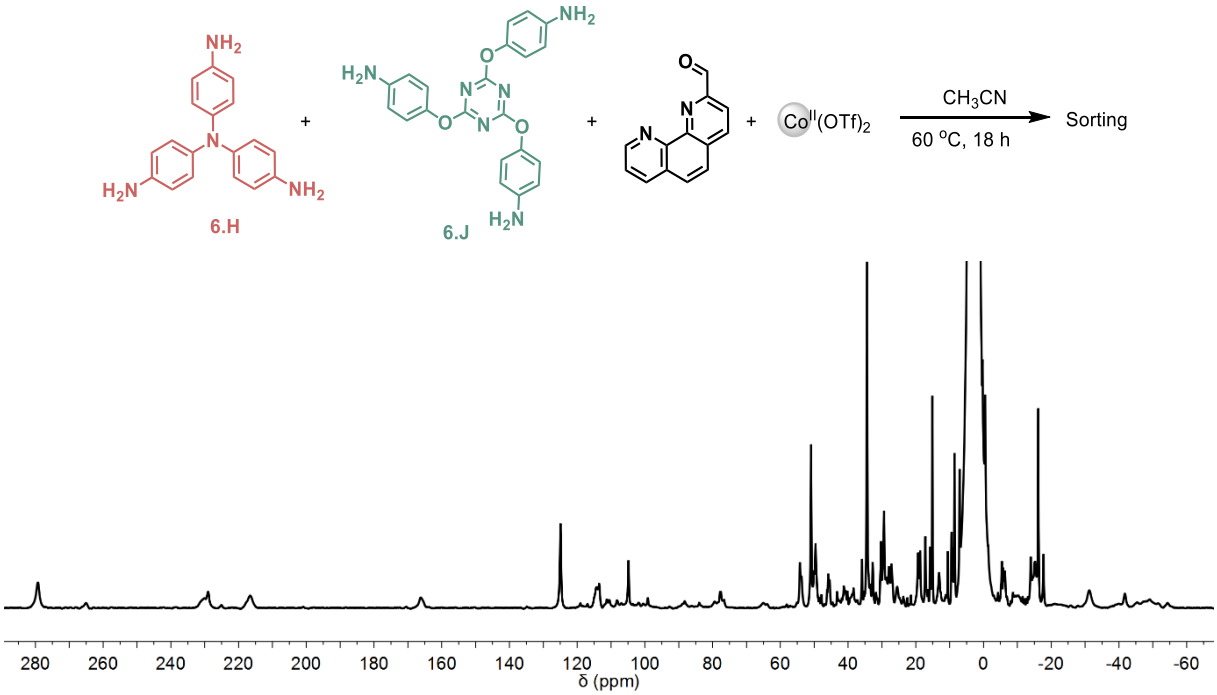


Figure 6.60 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of the mixture of cages generated when both **6.H** and **6.J** were used in the self-assembly process.

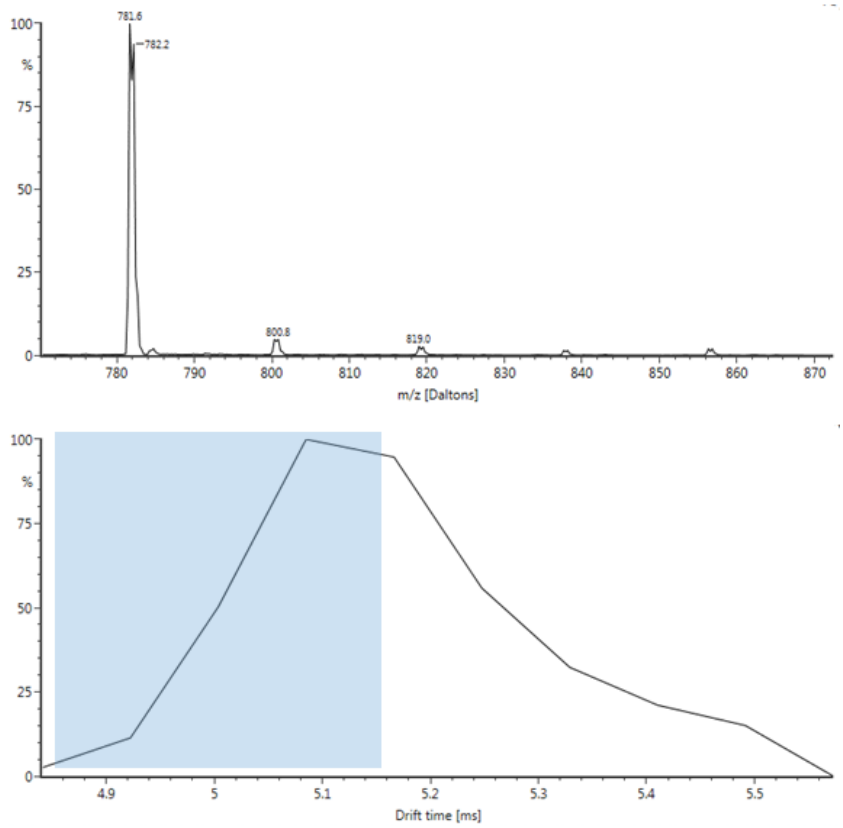


Figure 6.61 | IM-MS data for cages generated when both **6.H** and **6.J** were used in the self-assembly process, showing the z = +6 peaks (top). The species resolved with broad, asymmetric drift times, indicative of multiple species sizes in the assembly for this charge fragment set.

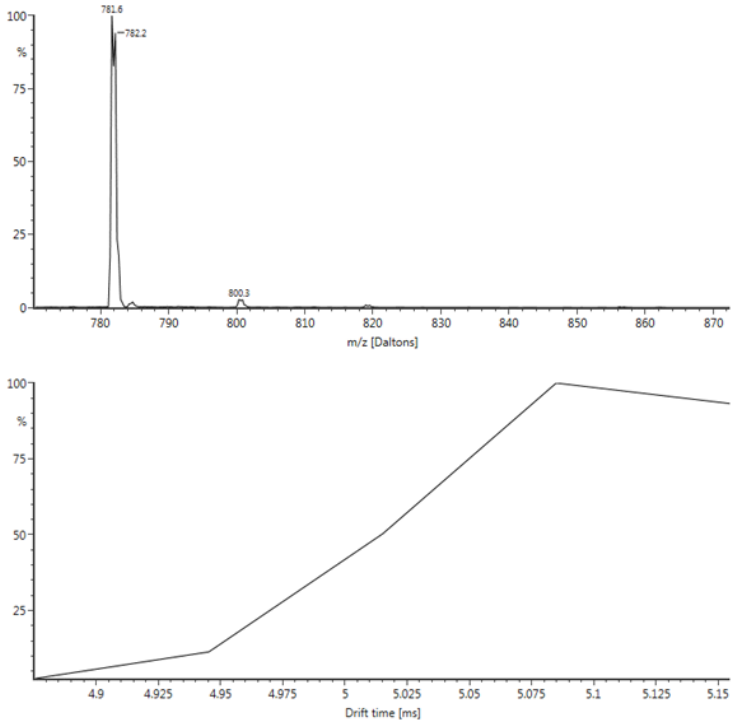


Figure 6.62 | When only the blue region in the drift time plot in Figure 6.61 was examined (i.e. the species with faster drift times), only peaks corresponding to species with a greater proportion of subcomponent **6.H** were observed.

Subcomponents 6.H and 6.K

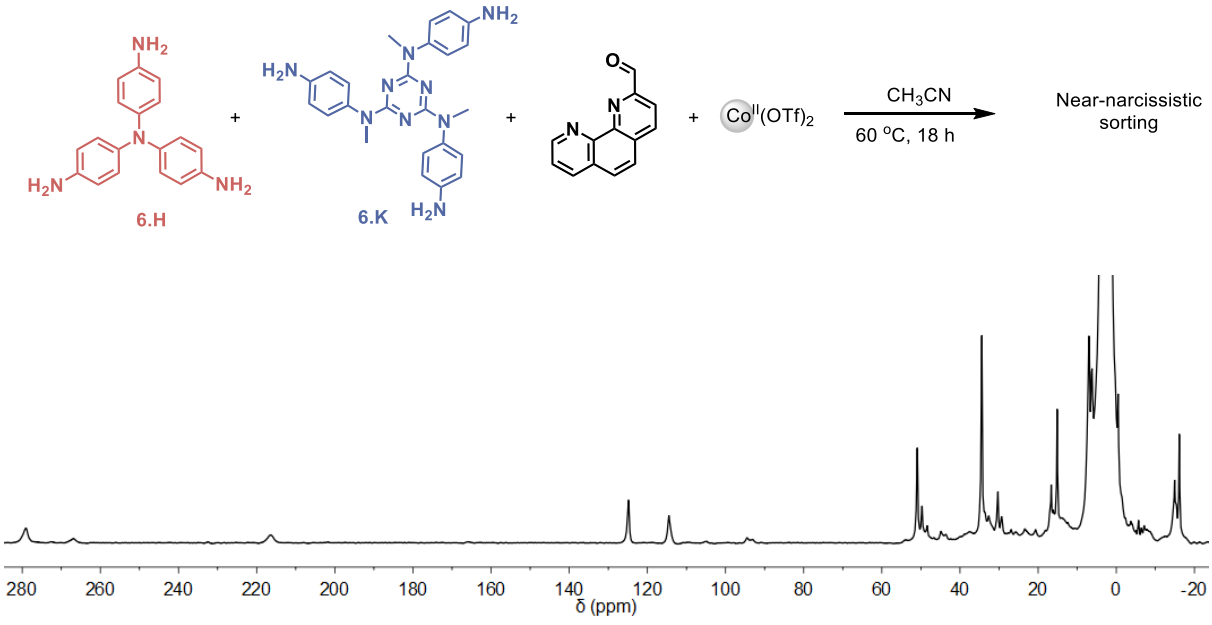


Figure 6.63 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of the mixture of cages generated when both **6.H** and **6.K** were used in the self-assembly process.

IM-MS data for cages generated when both **6.H** and **6.K** were used in the self-assembly process is given in Figure 6.13.

Subcomponents 6.I and 6.J

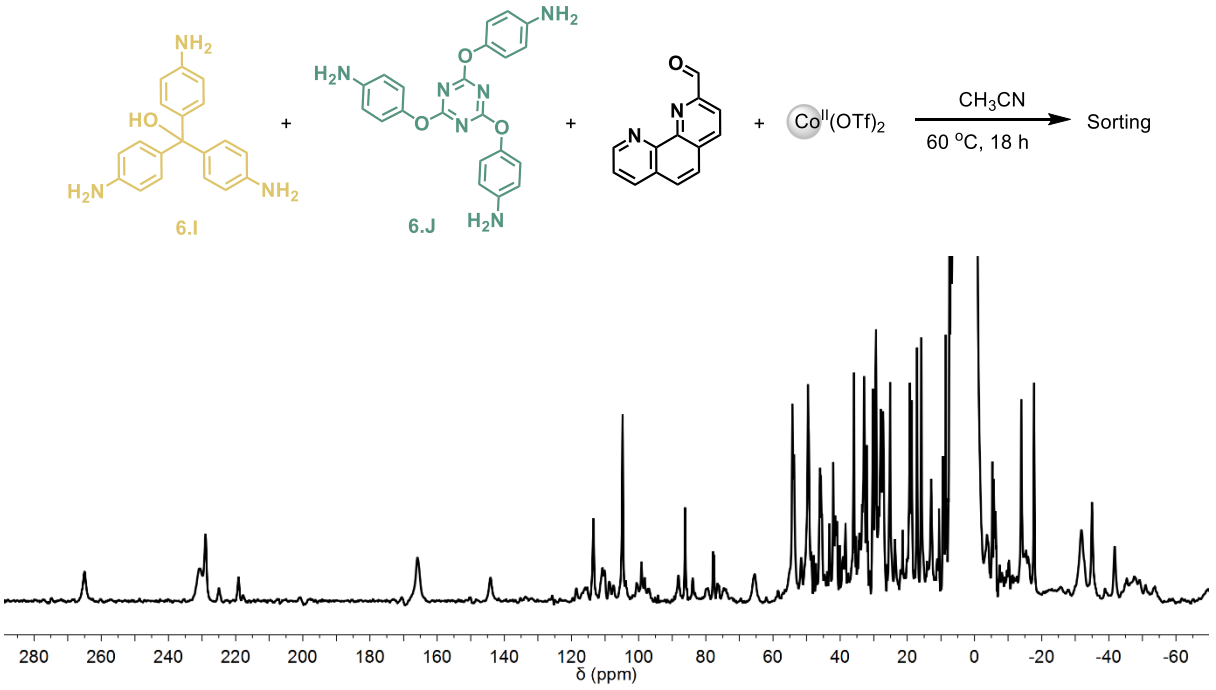


Figure 6.64 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of the mixture of cages generated when both **6.I** and **6.J** were used in the self-assembly process.

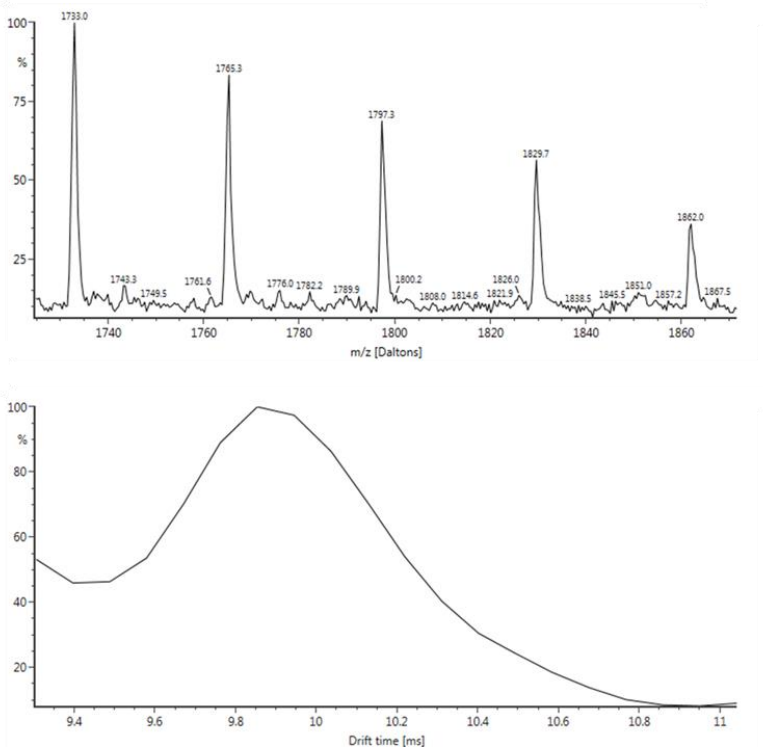


Figure 6.65 | IM-MS data for cages generated when both **6.I** and **6.J** were used in the self-assembly process, showing the z = +3 peaks (top). The species resolved with a single broad drift time, indicative of the octahedral geometry of all sorted possibilities.

Subcomponents 6.I and 6.K

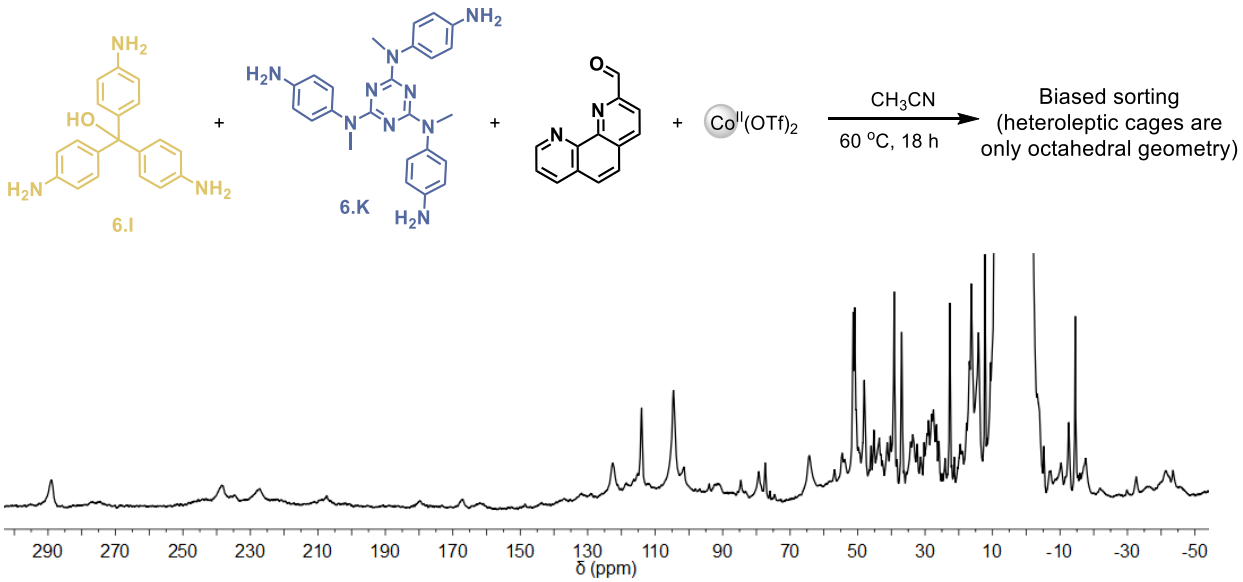


Figure 6.66 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of the mixture of cages generated when both **6.I** and **6.K** were used in the self-assembly process.

IM-MS data for cages generated when both **6.I** and **6.K** were used in the self-assembly process is given in Figure 6.14.

Subcomponents 6.J and 6.K

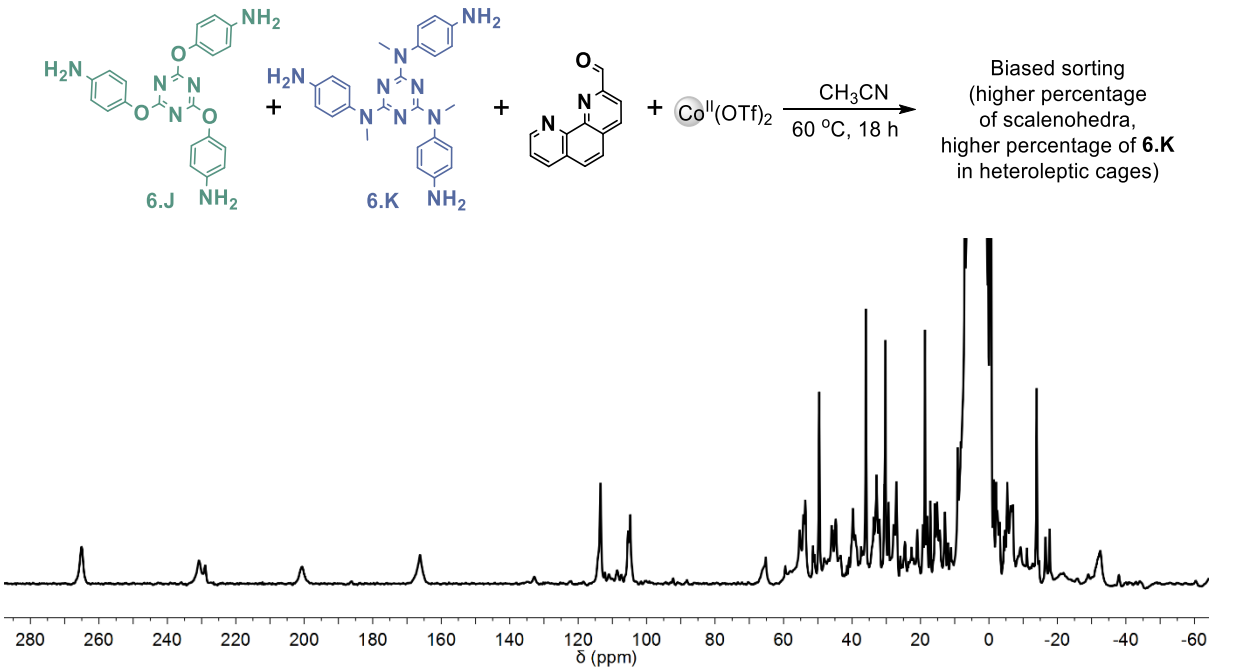
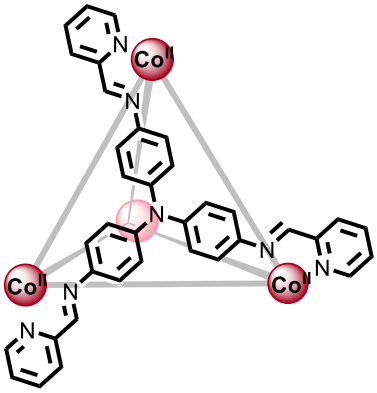


Figure 6.67 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of the mixture of cages generated when both **6.J** and **6.K** were used in the self-assembly process.

IM-MS data for cages generated when both **6.J** and **6.K** were used in the self-assembly process is given in Figure 6.15.

6.6.7 Synthesis and characterization of cages 6.9 and 6.10

Cage 6.9



Tris(4-aminophenyl)amine **6.H** (9.0 mg, 31.1 µmol, 1.0 eq.), picolinaldehyde (10.0 mg, 93.3 µmol, 3.0 eq.) and $Co(NTf_2)_2$ (21.5 mg, 31.1 µmol, 1.0 eq.) were combined in CH_3CN or CD_3CN (2.5 mL) in a sealed 5 mL reaction tube. The solution was stirred and heated at 70 °C for 18 h. A dark orange stock solution of cage **6.9** (3.11 mM) was obtained and was used without further purification.

^{1}H NMR (400 MHz, 298K, CD_3CN) δ 234.6 (s, 12H), 87.6 (s, 12H), 71.5 (s, 12H), 51.0 (s, 12H), 14.8 (s, 12H), –7.7 (s, 24H), –25.1 (s, 24H). **^{19}F NMR** (376 MHz, 298K, CD_3CN) δ –76.82. **LR-ESI-MS** [charge, calculated mass]: *m/z* = 1288.9 [**6.9**$(NTf_2)_5^{3+}$, 1288.0], 896.7 [**6.9**$(NTf_2)_4^{4+}$, 896.0], 661.3 [**6.9**$(NTf_2)_3^{5+}$, 660.8], 504.4 [**6.9**$(NTf_2)_2^{6+}$, 504.1], 392.4 [**6.9**$(NTf_2)^{7+}$, 392.1], 308.4 [**6.9**$^{8+}$, 308.1].

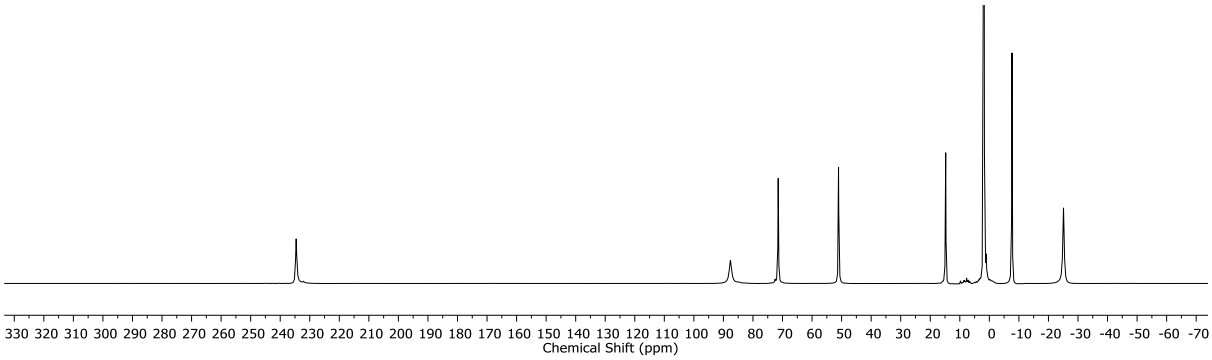


Figure 6.68 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of **6.9**.

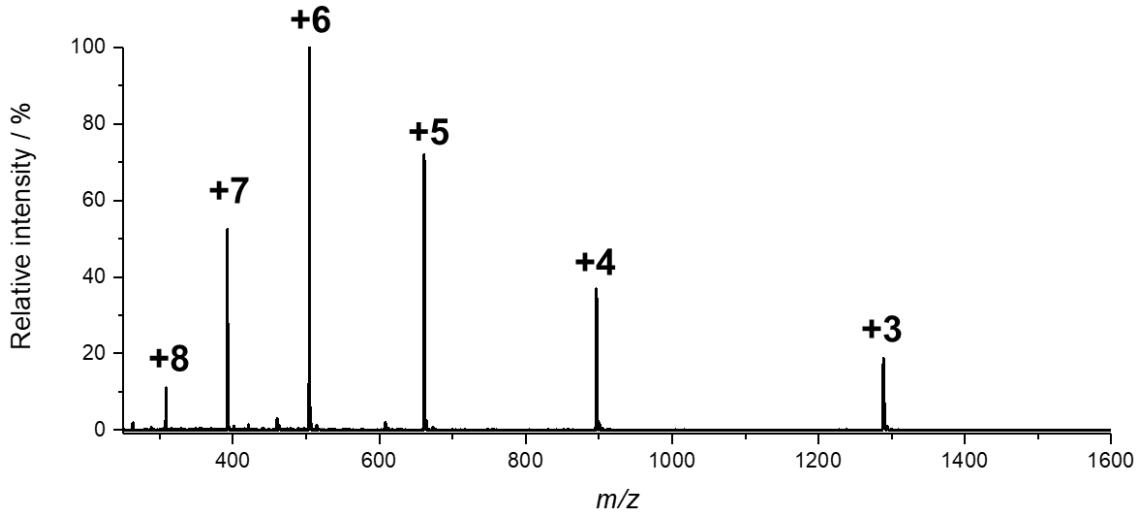


Figure 6.69 | LR-ESI-MS mass spectrum of **6.9**$(NTf_2)_8$.

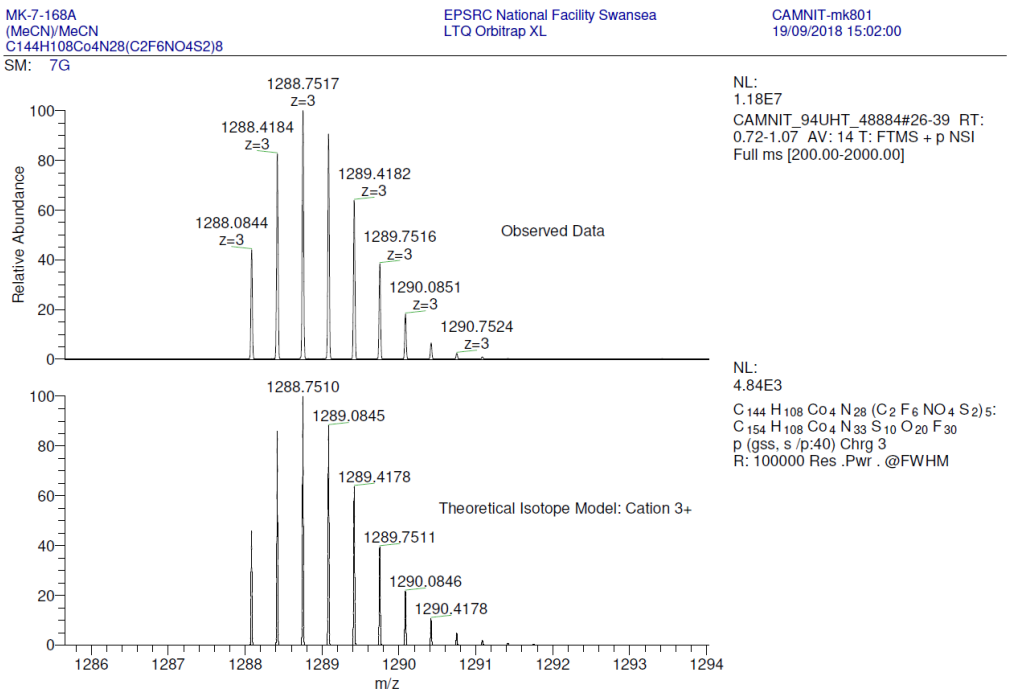
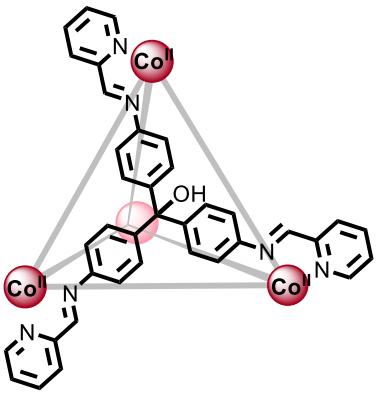


Figure 6.70 | HR-ESI-Mass spectrum of **6.9**$(NTf_2)_8$ showing the observed z = +3 charge, (top) compared to the theoretical isotope pattern (bottom).

Cage 6.10

With $Co(NTf_2)_2$



Pararosaniline **6.I** (9.0 mg, 31.1 µmol, 1.0 eq.), picolinaldehyde (10.0 mg, 93.3 µmol, 3.0 eq.) and $Co(NTf_2)_2$ (21.5 mg, 31.1 µmol, 1.0 eq.) were combined in CH_3CN or CD_3CN (2.5 mL) in a sealed 5 mL reaction tube. The solution was stirred and heated at 70 °C for 18 h. A dark red stock solution of cage **6.10** (3.11 mM) was obtained and was used without further purification.

A second minor set of signals was observed in the ^{1}H NMR spectrum which was attributed to the host-guest complex $CD_3CN\subset$**6.10**.

^{1}H NMR (400 MHz, 298K, CD_3CN) δ 240.1 (s, 12H), 236.4 (s, 12H, HG complex), 89.0 (s, 12H), 87.3 (s, 12H, HG complex), 73.6 (s, 24H), 52.0 (s, 24H), 16.2 (s, 24H), –4.666 (s, 48H), –23.8 (s, 48H). **^{19}F NMR** (376 MHz, 298K, CD_3CN) δ –79.56. **LR-ESI-MS** [charge, calculated mass]: *m/z* = 1308.9 [**6.10**$(NTf_2)_5^{3+}$, 1308.1], 911.7 [**6.10**$(NTf_2)_4^{4+}$, 911.0], 673.3 [**6.10**$(NTf_2)_3^{5+}$, 672.8], 514.4 [**6.10**$(NTf_2)_2^{6+}$, 514.1], 400.9 [**6.10**$(NTf_2)^{7+}$, 400.6], 315.9 [**6.10**$^{8+}$, 315.6].

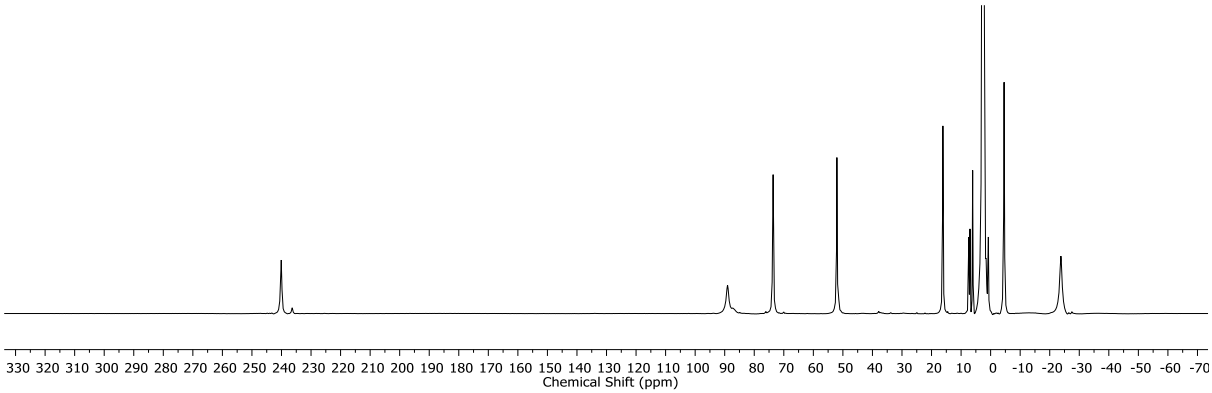


Figure 6.71 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of **6.10**$(NTf_2)_8$.

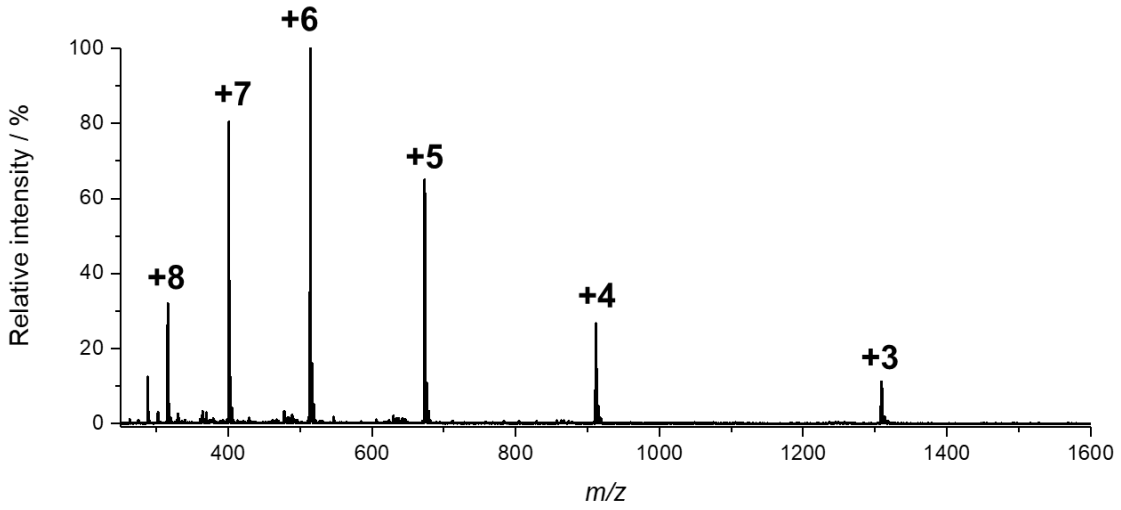


Figure 6.72 | LR-ESI-Mass spectrum of **6.10**$(NTf_2)_8$.

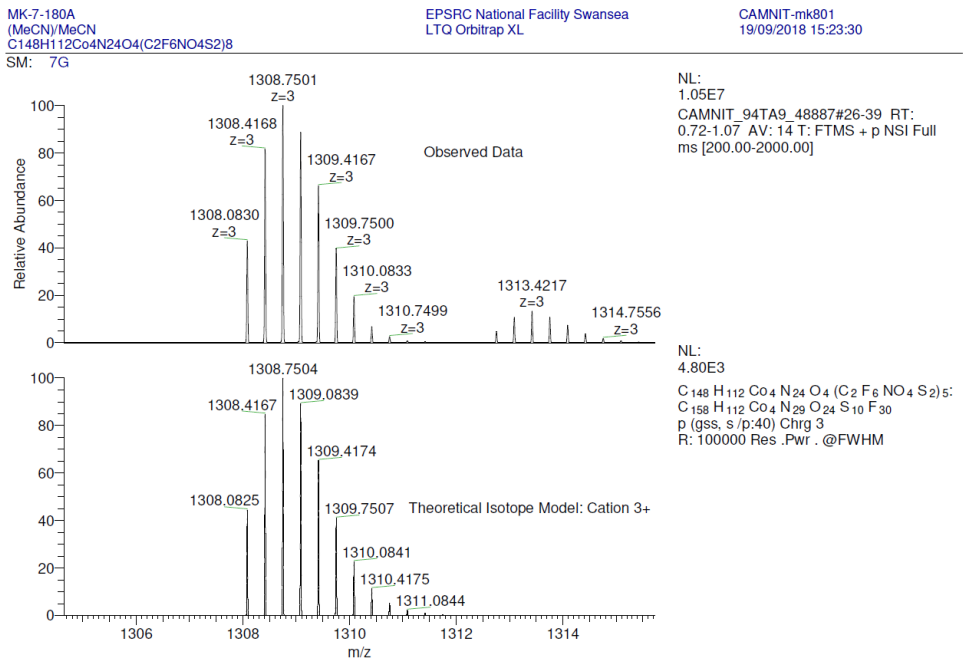
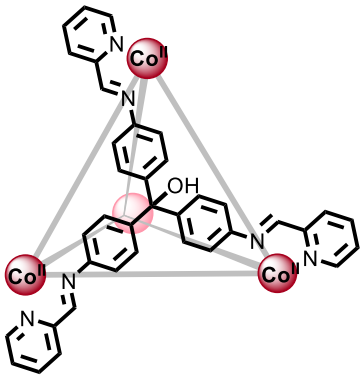


Figure 6.73 | HR-ESI-Mass spectrum of **6.10**$(NTf_2)_8$ showing the observed z = +3 charge, (top) compared to the theoretical isotope pattern (bottom).

With $Co(BF_4)_2$



Pararosaniline **6.I** (9.0 mg, 31.1 µmol, 1.0 eq.), picolinaldehyde (10.0 mg, 93.3 µmol, 3.0 eq.) and $Co(BF_4)_2$ (10.6 mg, 31.1 µmol, 1.0 eq.) were combined in CH_3CN or CD_3CN (2.5 mL) in a sealed 5 mL reaction tube. The solution was stirred and heated at 70 °C for 18 h. A dark red stock solution of cage **6.10** (3.11 mM) was obtained and was used without further purification.

Three set of signals were observed in the ^{1}H NMR which were attributed to the empty cage, the host-guest complex $CD_3CN\subset$**6.10** and the host-guest complex $BF_4^-\subset$**6.10**.

^{1}H NMR (400 MHz, 298K, CD_3CN) δ 239.7 (s, 1H, empty cage), 238.5 (s, 1H, $BF_4^-\subset$**6.10**), 236.1 (s, 1H, $CD_3CN\subset$**6.10**), 88.6 (bs, 1H, empty cage), 86.7 (bs, 1H, $CD_3CN\subset$**6.10**), 86.2 (bs, 1H, $BF_4^-\subset$**6.10**), 74.0 – 72.9 (m, 3H), 51.8 (s, 3H), 16.4 (s, 3H), –4.5 (s, 6H), –23.6 (bs, 2H, empty cage and $CD_3CN\subset$**6.10**), –25.7 (bs, 2H, $BF_4^-\subset$**6.10**). **^{19}F NMR** (376 MHz, 298K, CD_3CN) δ –143.45 (s, free BF_4^-), –200.02 (s, $BF_4^-\subset$**6.10**). **LR-ESI-MS** [charge, calculated mass]: *m/z* = 1523.4 [**6.10**$(BF_4)_6^{2+}$, 1523.24], 986.7 [**6.10**$(BF_4)_5^{3+}$, 986.5], 718.3 [**6.10**$(BF_4)_4^{4+}$, 718.1], 557.2 [**6.10**$(BF_4)_3^{5+}$, 557.1], 449.8 [**6.10**$(BF_4)_2^{6+}$, 449.7], 373.2 [**6.10**$(BF_4)^{7+}$, 373.1], 315.1 [**6.10**$^{8+}$, 315.6].

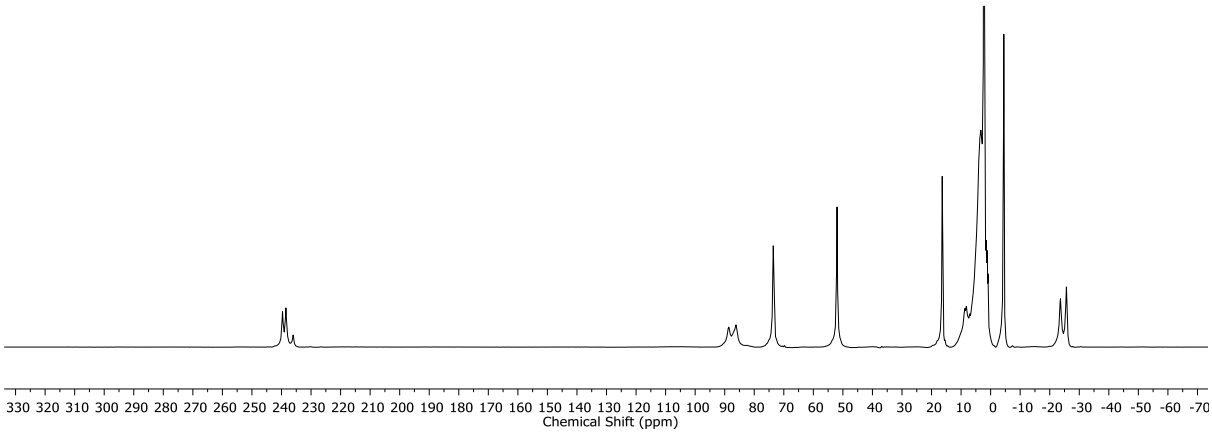


Figure 6.74 | Wide sweep ^{1}H NMR spectrum (400 MHz, 298 K, CD_3CN) of **6.10**$(BF_4)_8$.

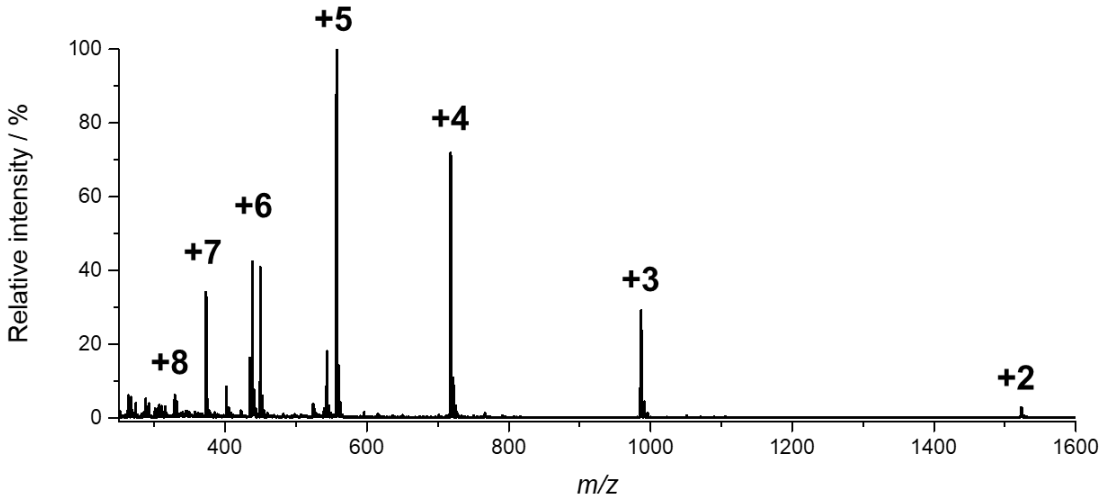


Figure 6.75 | LR-ESI-MS mass spectrum of **6.10**$(BF_4)_8$.

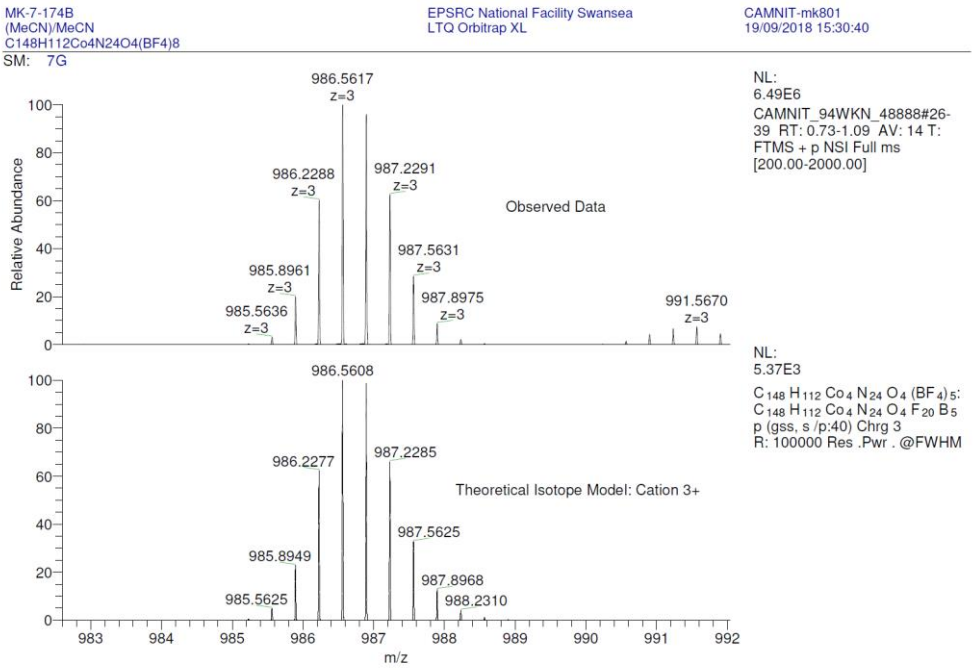


Figure 6.76 | HR-ESI-MS mass spectrum of **6.10**$(BF_4)_8$ showing the observed z = +3 charge, (top) compared to the theoretical isotope pattern (bottom).

Crystal structure of 6.9 and 6.10

Crystals of **6.9** suitable for single crystal X-ray diffraction analysis were obtained by Rana Bilbeisi through vapour diffusion of diethyl ether into an acetonitrile solution of **6.9** to which $TBAClO_4$ was added. Crystals of **6.10** suitable for single crystal X-ray diffraction analysis were obtained by Rana Bilbeisi through vapour diffusion of a solution of 50:50 (v/v) diethyl ether:ethyl acetate into an acetonitrile solution of **6.10**$(ClO_4^-)_8$. Data sets were collected by Jack Clegg.

6.6.8 Host-guest chemistry in 6.10

Binding of acetonitrile

A solution of cage **6.10** (3 mM) was prepared in CH_3CN according to the procedure described in section 6.6.7. A capillary of CD_3CN was added to the NMR tube containing the cage in the non-deuterated solvent in order to allow locking and a first ^1H NMR spectrum was collected. The sample was dried, dissolved in CD_3CN and left to equilibrate at 70 °C for 4 days after which the ^1H NMR spectrum was collected again.

Binding of ClO_4^-

$TBAClO_4$ was titrated into a solution of cage **6.10** (1.0 mM). The samples were left to equilibrate for 24 h at 70 °C after each addition of $TBAClO_4$ before collecting the spectrum.

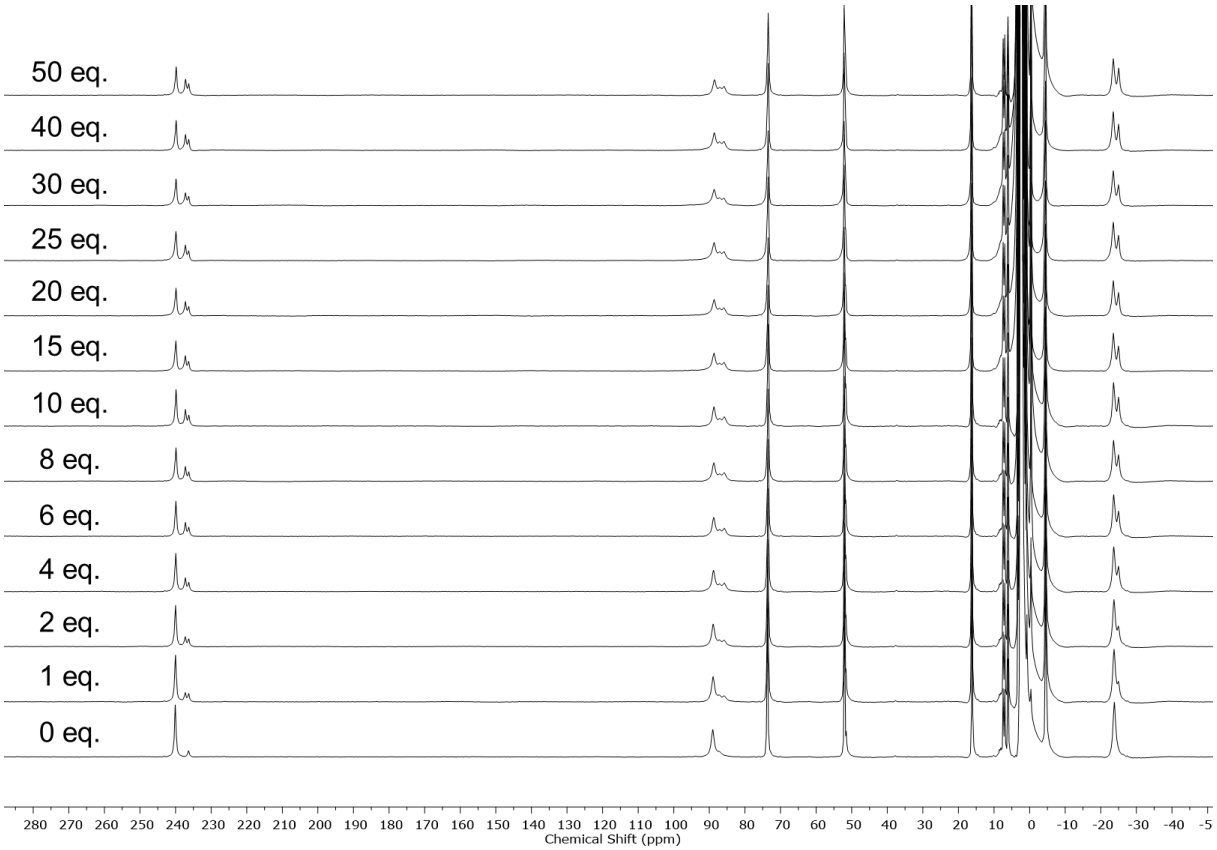


Figure 6.77 | Wide sweep ^1H NMR spectra (400 MHz, 298 K, CD_3CN) of cage **6.10** after titration of $TBAClO_4$ (equivalents of anions labelled on the spectra).

Binding of BF_4^-

$TBABF_4$ was titrated into a solution of cage **6.10** (1.0 mM). The samples were left to equilibrate for 24 h at 70 °C after each addition of $TBABF_4$ before collecting the spectrum.

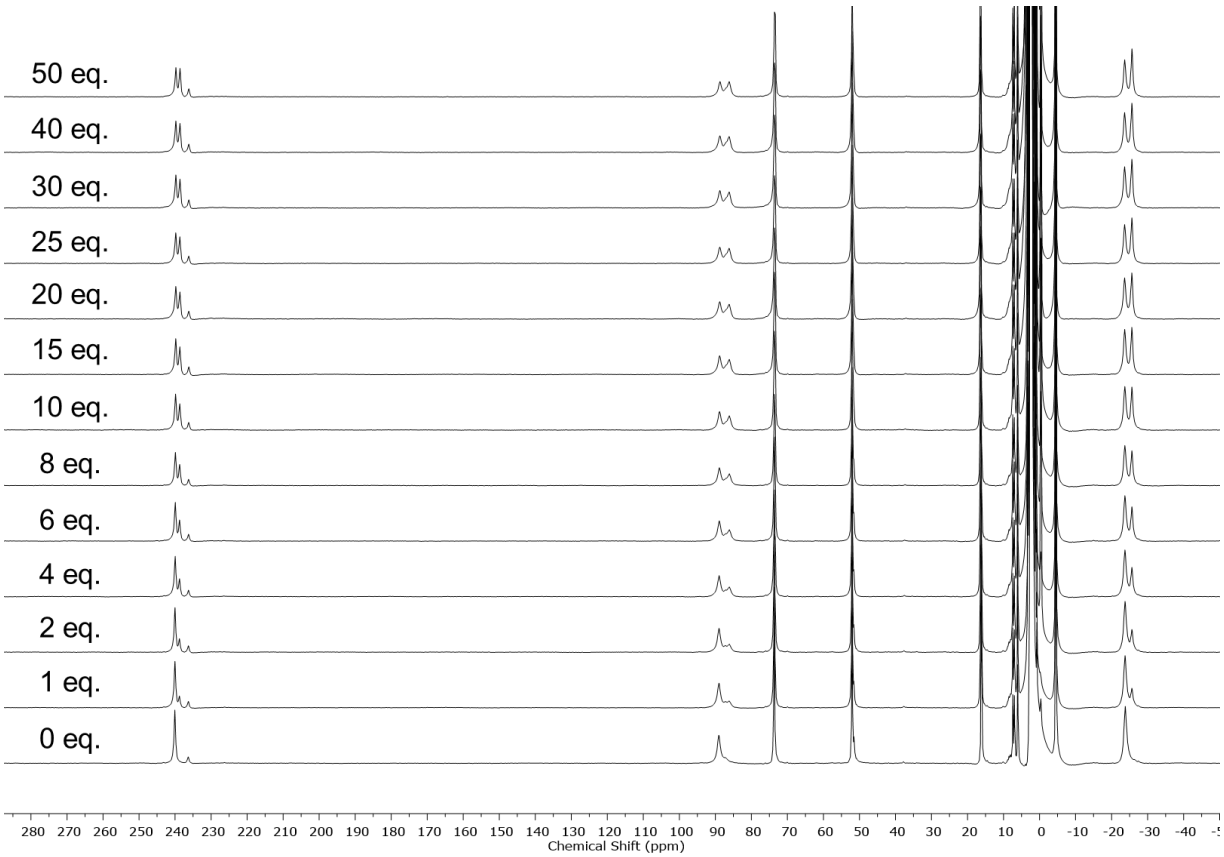


Figure 6.78 | Wide sweep ^1H NMR spectra (400 MHz, 298 K, CD_3CN) of cage **6.10** after titration of $TBABF_4$ (equivalents of anions labelled on the spectra).

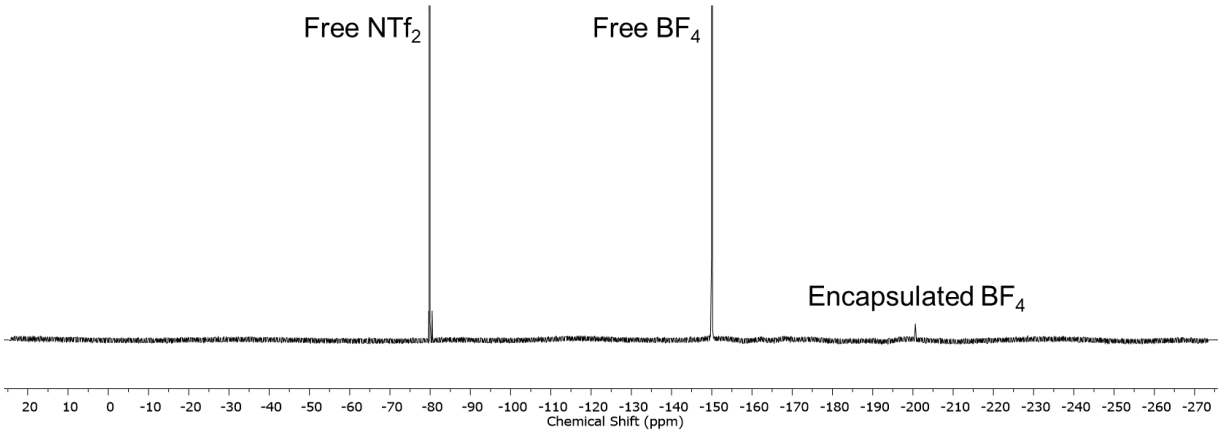


Figure 6.79 | ^{19}F spectrum (376 MHz, 298 K, CD_3CN) of cage **6.10** after addition of 50 eq. of $TBABF_4$.

Binding of I^-

TBAI was titrated into a solution of cage **6.10** (1.0 mM). The samples were left to equilibrate for 24 h at 70 °C after each addition of TBAI before collecting the spectrum.

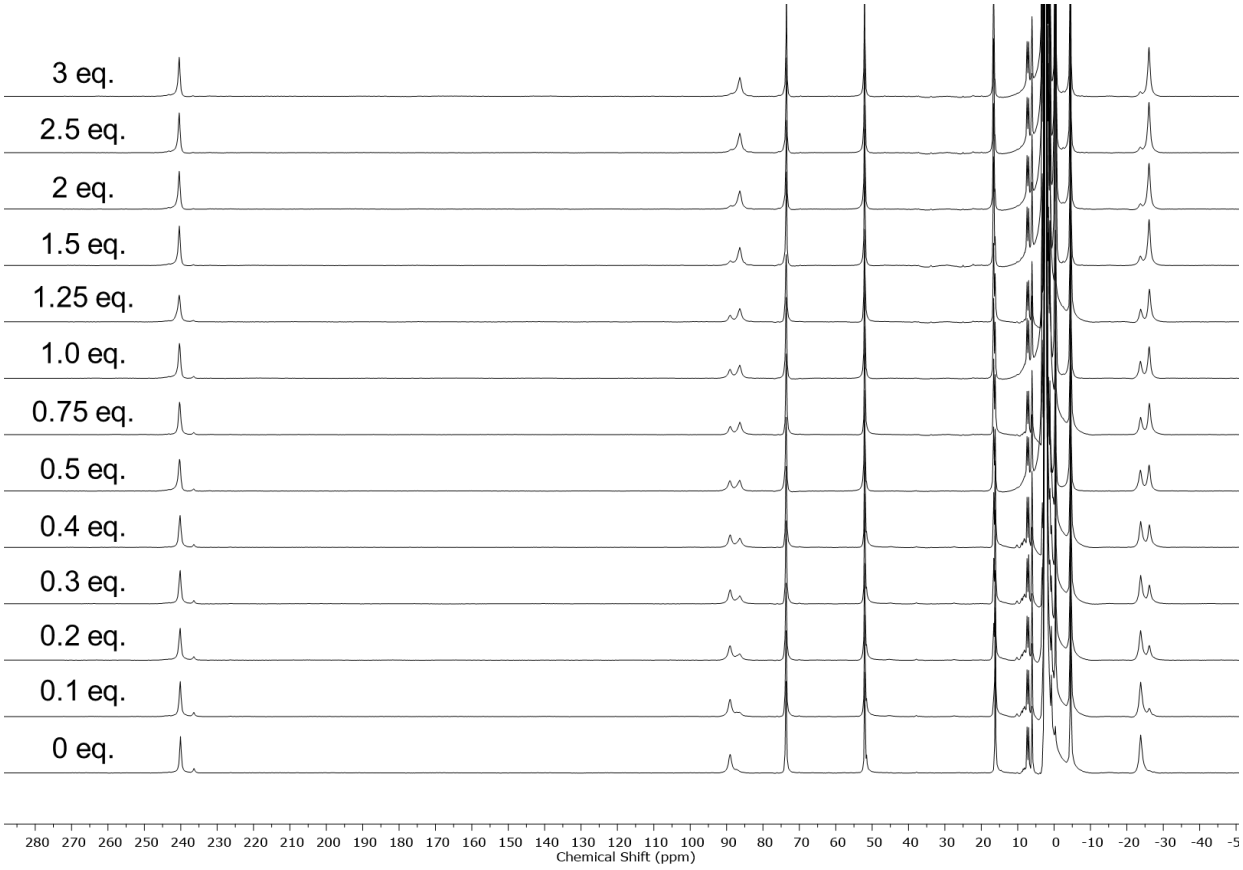


Figure 6.80 | Wide sweep ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) of cage **6.10** after titration of TBAI (equivalents of anions labelled on the spectra)

Binding of Br^-

TBABr was titrated into a solution of cage **6.10** (1.0 mM). The samples were left to equilibrate for 24 h at 70 °C after each addition of TBABr before collecting the spectrum.

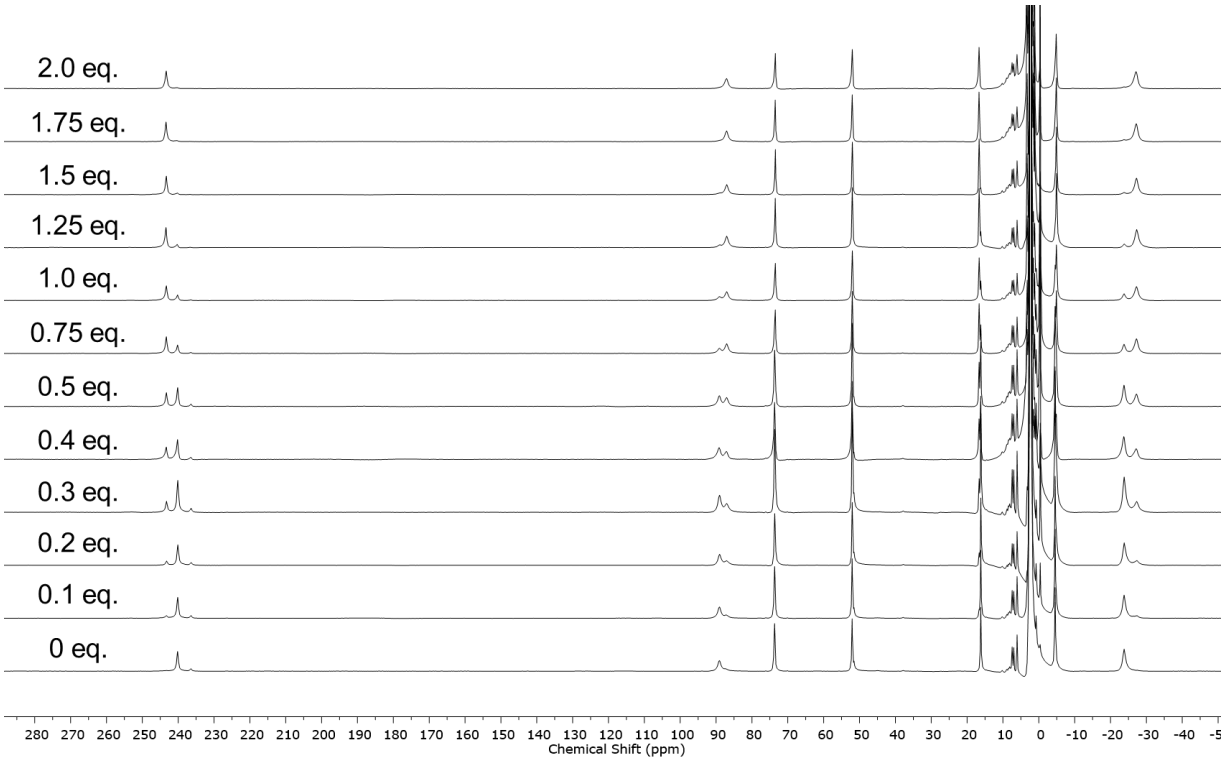


Figure 6.81 | Wide sweep ^{1}H NMR spectra (400 MHz, 298 K, CD_3CN) of cage **6.10** after titration of TBABr (equivalents of anions labelled on the spectra)

6.6.9 Mathematical models for titration curves

Definition of [HG], [HS] and [H]

Cage **6.10** was found to undergo 1:1 binding with two types of guests: anions such as Br^-, Cl^-, BF_4^-, and ClO_4^- and acetonitrile.

$$H + G \rightleftharpoons HG \qquad\qquad H + S \rightleftharpoons HS$$

Each of these binding events can be expressed by the following equilibrium constants:

$$K_{HG} = \frac{[HG]}{[H][G]} \qquad \text{(Equation 6.10)}$$

$$K_{HS} = \frac{[HS]}{[H][S]} \qquad \text{(Equation 6.11)}$$

where [*H*], [*G*], [*S*] are the concentrations of the unbound species and [*HG*] and [*HS*] represent the bound species. Given that the solvent concentration (~20,000 mM) is orders of magnitude greater than that of either the host or guest (~1 mM), the change in concentration upon encapsulation of solvent is very small, so that $[S_{free}] \approx [S_{total}]$.

From mass balance:

$$[G_t] = [G] + [HG] \qquad \text{(Equation 6.12)}$$

$$[H_t] = [H] + [HG] + [HS] = [H] + [HG] + K_{HS}[H][S_t] = (1 + K_{HS}[S_t])[H] + [HG] \qquad \text{(Equation 6.13)}$$

The free host concentration *H* is thus:

$$[H] = \frac{[H_t] - [HG]}{(1 + K_{HS}[S_t])} \qquad \text{(Equation 6.14)}$$

Substitution into the equilibrium constant expressions gives:

$$K_{HG} = \frac{[HG]}{[H][G]} = \frac{(1 + K_{HS}[S_t])[HG]}{([H_t] - [HG])([G_t] - [HG])} \qquad \text{(Equation 6.15)}$$

Rearranging terms, followed by division with K_{HG} gives the following quadratic equation:

$$[HG]^2 - \left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} \right)[HG] + [H_t][G_t] = 0 \qquad \text{(Equation 6.16)}$$

Solving for the host-guest complex [*HG*] using the quadratic formula gives the following hyperbola:

$$[HG] = \frac{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} \right) - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} \right)^2 - 4[H_t][G_t]}}{2} \qquad \text{(Equation 6.7)}$$

The root with the negative radical rather than the positive radical was used in order to satisfy boundary conditions, namely that at $[G_t] = 0$, [*HG*] must equal zero.

Likewise, an analytical function can be written for free host [*H*] as a function of [G_t]. According to equation 6.14, *H* can be expressed in terms of *HG*. Substituting the host-guest complex concentration defined by equation 6.7 gives:

$$[H] = \frac{[H_t] - \left(\frac{1}{2} \left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} \right)^2 - 4[H_t][G_t]} \right) \right)}{(1 + K_{HS}[S_t])}$$

(Equation 6.8)

Likewise, an analytical function can be written for [*HS*] as a function of [G_t]. Equation 6.13 can be rewritten to give bound host-solvent complex as:

$$[HS] = ([H_t] - [HG]) \left(\frac{K_{HS}[S_t]}{1 + K_{HS}[S_t]} \right)$$

(Equation 6.17)

Substituting the definition of [*HG*] in equation 6.7 gives:

$$[HS] = \frac{K_{HS}[S_t]}{1 + K_{HS}[S_t]} \left([H_t] - \frac{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} \right) - \sqrt{\left([H_t] + [G_t] + \frac{(1 + K_{HS}[S_t])}{K_{HG}} \right)^2 - 4[H_t][G_t]}}{2} \right)$$

(Equation 6.9)

Equation 6.7, 6.8 and 6.9 were used for the case of TBABr and TBAI. A scaling factor C was added to Equation 6.7, 6.8 and 6.9 to account for the contribution of TBA^+. Furthermore, giving that [*HS*] did not change required elimination of the terms related to the otherwise competitive binding of solvent, namely 1 + K_{HS} [*S*] term with concomitant addition of the term –[*HS*] to keep the proper mass balance. Finally, although [*HS*] is essentially constant throughout the titration with larger guests, small variations were modelled as a straight line, giving Equation 6.10, 6.11 and 6.12 described in Section 6.4.2.

The Merit Function χ^2

The following work was carried out solely by John Thoburn.

For Gaussian distribution of measurements with individual standard deviations σ_i, the probability that the model parameters fit the N data points is

$$P \propto \prod_{i=1}^{N} e^{-\left(y_i^{obs} - y_i^{calc}\right)^2 / 2\sigma_i^2}$$

(Equation 6.18)

where y_i^{obs} is the range of the *ith* data point and y_i^{calc} is the range calculated for the corresponding *ith* data point according to equations 6.7, 6.8, and 6.9. Taking the log of the probability allows one to work with sums rather than products, which are computationally more readily handled than exponents:

$$\ln(P) \propto -\sum_{i=1}^{N} \frac{\left(y_i^{obs} - y_i^{calc}\right)^2}{\sigma_i^2}$$

(Equation 6.19)

The quantity to be maximized, ln(*P*), is equivalent to minimizing –ln(*P*)

$$-\ln(P) \propto \sum_{i=1}^{N} \frac{\left(y_i^{obs} - y_i^{calc}\right)^2}{\sigma_i^2}$$

(Equation 6.20)

Thus, a "merit" function, or fitting function, χ^2 was defined, which is the function to be minimized during the best fit determination of the desired parameters.

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(y_i^{obs} - y_i^{calc}\right)^2}{\sigma_i^2}$$

(Equation 6.21)

Some data points were more reliable than others and therefore each data point's contribution to the fit was weighted by $1/\sigma_i^2$, where σ_i^2 is the variance. The term y_i^{calc} is a function of the variable [*Gt*] and the binding constants K_{HG} and K_{HS}, which are the parameters to be fitted during the minimization process. Because all three curves are dependent on the parameters K_{HG} and K_{HS}, the best fit is obtained by fitting all three functions simultaneously rather than individually. Thus, we define the merit function:

$$\chi^2 = \chi_{HG}^2 + \chi_H^2 + \chi_{HS}^2$$

(Equation 6.22)

A Mathematica program was written to implement this. Each data points is to be fitted to only one of the equations 6.7, 6.8, or 6.9, that is to only one of the χ^2 terms in equation 6.21. To sort the data to the appropriate term, each data point was assigned an index (1, 2, or 3), corresponding to equations 6.7, 6.8, or 6.9. Thus, the N data points to be fitted were represented

as an N by 3 matrix, where each row consisted of {$index_i$, x_i, y_i}. Then the Kronecker delta was used to associate individual data points to the appropriate fitting function:

$$\chi^2 = \delta_{ij}\chi^2_{HG} + \delta_{ij}\chi^2_{H} + \delta_{ij}\chi^2_{HS} \qquad \text{(Equation 6.23)}$$

where $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$.

Minimization Routine and Error Propagation

The binding constants were calculated by minimizing merit function (Equation 6.23) with respect to the fitted parameters, K_{HG} and K_{HS}:

$$\frac{\partial \chi^2}{\partial K_{HG}} = 0 \qquad \text{(Equation 6.24a)}$$

$$\frac{\partial \chi^2}{\partial K_{HS}} = 0 \qquad \text{(Equation 6.24b)}$$

The roots to equations 6.27 were found numerically using the Newton-Raphson method as implemented Mathematica's built-in NonlinearModelFit function. In this iterative method a new guess at the root x_{n+1} is derived from an initial guess (x_n), the value of the function at x_n and its derivative:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad \text{(Equation 6.25)}$$

In some cases better convergence was achieved with the Levenberg-Marquardt minimization procedure. The standard deviations ("errors") in the fitted parameters were determined by the usual propagation of errors:

$$dK_{HG} = \sqrt{\left(\frac{\partial K_{HG}}{\partial HG}\right)^2_{HS} dHG^2} \qquad \text{(Equation 6.26a)}$$

$$dK_{HS} = \sqrt{\left(\frac{\partial K_{HS}}{\partial HS}\right)^2_{HG} dHS^2} \qquad \text{(Equation 6.26b)}$$

where *dHG* and *dHS* is the uncertainty of [HG] and [HS], respectively, as estimated from their residuals.

The Mathematica program will be made freely available as supplementary information upon publication of this work.

6.6.10 Sorting experiments for cages 6.9 and 6.10

Untemplated library of heteroleptic cages

Tris(4-aminophenyl)amine **6.H** (4.5 mg, 15.6 µmol, 1.0 eq.), pararosaniline **6.I** (4.75 mg, 15.6 µmol, 1.0 eq.), picolinaldehyde (10.0 mg, 93.3 µmol, 6.0 eq.) and $Co(NTf_2)_2$ (21.5 mg, 31.1 µmol, 2.0 eq.) were combined in CH_3CN or CD_3CN (2.5 mL) in a sealed 5 mL reaction tube. The solution was stirred and heated at 70 °C for 18 h. A dark red stock solution of the untemplated library of cages (3.11 mM) was obtained and was used without further purification.

Templation with Br^- or I^-

TBAI or TBABr (2 eq.) were added to a solution of the untemplated mixture (3.11 mM). The solutions were stirred and heated at 70 °C for 72 h. A dark red solution of the templated library of cages (3.11 mM) was obtained and was used without further purification.

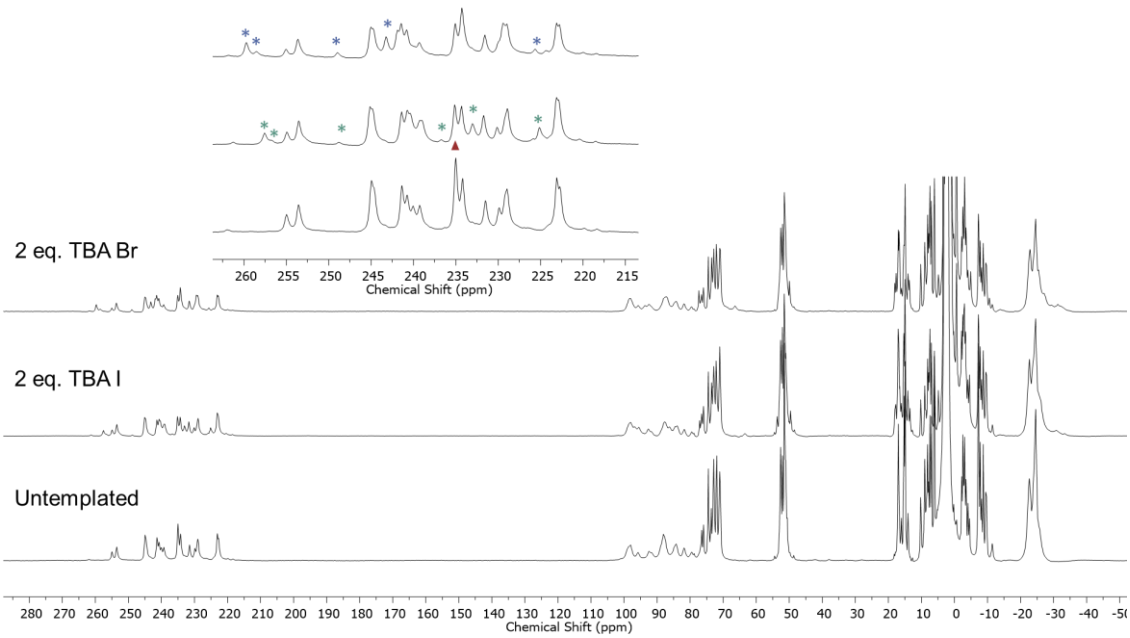


Figure 6.82 | Wide sweep ^1H NMR spectrum (400 MHz, 298 K, CD_3CN) of the untemplated library of cages (bottom) and the templated library by I^- (middle) and Br^- (top). The new peaks corresponding to the encapsulated guests within the library are marked by blue stars and green stars. The peak corresponding to **6.9** is marked by a red triangle.

Templation with ClO_4^- or BF_4^-

Tris(4-aminophenyl)amine **6.H** (4.5 mg, 15.60 µmol, 1.0 eq.), pararosaniline **6.I** (4.75 mg, 15.60 µmol, 1.0 eq.), picolinaldehyde (10.0 mg, 93.30 µmol, 6.0 eq.), $Co(NTf_2)_2$ (21.5 mg, 31.10 µmol, 2.0 eq.) and TBA ClO_4 or TBA BF_4 (7.80 µmol, 0.5 eq.) were combined in CH_3CN or CD_3CN (2.5 mL) in a sealed 5 mL reaction tube. The solutions were stirred and heated at 70 °C for 72 h. A dark red stock solution of the templated library of cages (3.11 mM) was obtained and was used without further purification.

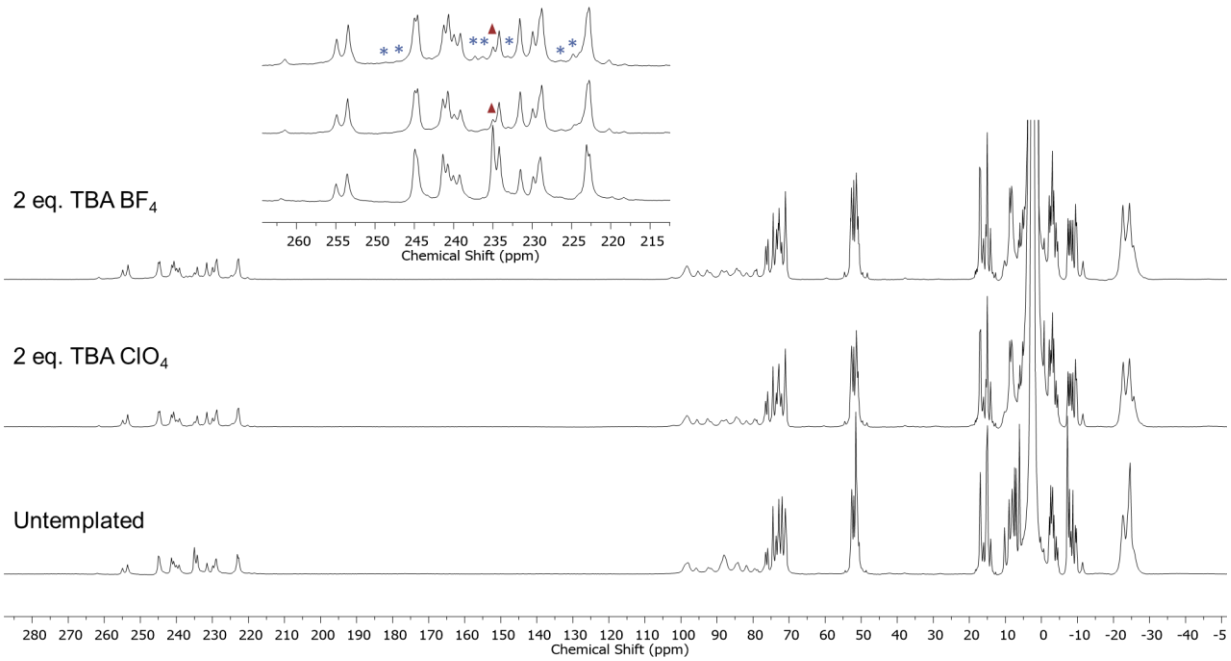


Figure 6.83 | Wide sweep ^1H NMR spectrum (400 MHz, 298 K, CD_3CN) of the untemplated library of cages (bottom) and the templated library by BF_4^- (middle) and ClO_4^- (top). The new peaks corresponding to the encapsulated guests within the library are marked by blue stars. The peak corresponding to **6.9** is marked by a red triangle.

6.7 References

[1] M. M. Safont-Sempere, G. Fernández, F. Würthner, *Chem. Rev.* **2011**, *111*, 5784-5814.

[2] D. Beaudoin, F. Rominger, M. Mastalerz, *Angew. Chem. Int. Ed.* **2017**, *56*, 1244-1248.

[3] Y. Tidhar, H. Weissman, D. Tworowski, B. Rybtchinski, *Chem. - Eur. J.* **2014**, *20*, 10332-10342.

[4] O. Š. Miljanić, *Chem* **2017**, *2*, 502-524.

[5] M. De Poli, W. Zawodny, O. Quinonero, M. Lorch, S. J. Webb, J. Clayden, *Science* **2016**, *352*, 575-580.

[6] I. V. Kolesnichenko, E. V. Anslyn, *Chem. Soc. Rev.* **2017**, *46*, 2385-2390.

[7] F. Otis, M. Auger, N. Voyer, *Acc. Chem. Res.* **2013**, *46*, 2934-2943.

[8] D. van der Zwaag, T. F. A. de Greef, E. W. Meijer, *Angew. Chem. Int. Ed.* **2015**, *54*, 8334-8336.

[9] S. Iamsaard, S. J. Aßhoff, B. Matt, T. Kudernac, J. J. L. M. Cornelissen, S. P. Fletcher, N. Katsonis, *Nat. Chem.* **2014**, *6*, 229-235.

[10] S. K. Samanta, D. Moncelet, V. Briken, L. Isaacs, *J. Am. Chem. Soc.* **2016**, *138*, 14488-14496.

[11] J. R. Kumpfer, S. J. Rowan, *ACS Macro Lett.* **2012**, *1*, 882-887.

[12] N. Avakyan, A. A. Greschner, F. Aldaye, C. J. Serpell, V. Toader, A. Petitjean, H. F. Sleiman, *Nat. Chem.* **2016**, *8*, 368-376.

[13] Z. Liu, G. Liu, Y. Wu, D. Cao, J. Sun, S. T. Schneebeli, M. S. Nassar, C. A. Mirkin, J. F. Stoddart, *J. Am. Chem. Soc.* **2014**, *136*, 16651-16660.

[14] N. Sinha, T. T. Y. Tan, E. Peris, F. E. Hahn, *Angew. Chem. Int. Ed.* **2017**, *56*, 7393-7397.

[15] J. E. M. Lewis, P. D. Beer, S. J. Loeb, S. M. Goldup, *Chem. Soc. Rev.* **2017**, *46*, 2577-2591.

[16] M. Obana, T. Fukino, T. Hikima, T. Aida, *J. Am. Chem. Soc.* **2016**, *138*, 9246-9250.

[17] W. M. Bloch, G. H. Clever, *Chem. Commun.* **2017**, *53*, 8506-8516.

[18] J. W. Sadownik, E. Mattia, P. Nowak, S. Otto, *Nat. Chem.* **2016**, *8*, 264-269.

[19] L. A. Ingerman, M. L. Waters, *J. Org. Chem.* **2009**, *74*, 111-117.

[20] M. Lal Saha, M. Schmittel, *Org. Biomol. Chem.* **2012**, *10*, 4651-4684.

[21] C. A. Schalley, *Analytical methods in supramolecular chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, **2007**.

[22] M. Kieffer, B. S. Pilgrim, T. K. Ronson, D. A. Roberts, M. Aleksanyan, J. R. Nitschke, *J. Am. Chem. Soc.* **2016**, *138*, 6813-6821.

[23] F. J. Rizzuto, M. Kieffer, J. R. Nitschke, *Chem. Sci.* **2018**, *9*, 1925-1930.

[24] M. Fujita, S. Nagao, K. Ogura, *J. Am. Chem. Soc.* **1995**, *117*, 1649-1650.

[25] K.-C. Sham, S.-M. Yiu, H.-L. Kwong, *Inorg. Chem.* **2013**, *52*, 5648-5650.

[26] E. Orhan, A. Garci, B. Therrien, *Inorg. Chim. Acta* **2015**, *438*, 5-9.

[27] K. Mahata, P. D. Frischmann, F. Würthner, *J. Am. Chem. Soc.* **2013**, *135*, 15656-15661.

[28] L.-L. Yan, C.-H. Tan, G.-L. Zhang, L.-P. Zhou, J.-C. Bünzli, Q.-F. Sun, *J. Am. Chem. Soc.* **2015**, *137*, 8550-8555.

[29] D.-H. Ren, D. Qiu, C.-Y. Pang, Z. Li, Z.-G. Gu, *Chem. Commun.* **2015**, *51*, 788-791.

[30] X. Jing, C. He, Y. Yang, C. Duan, *J. Am. Chem. Soc.* **2015**, *137*, 3967-3974.

[31] L. Pauling, *J. Am. Chem. Soc.* **1931**, *53*, 1367-1400.

[32] A. Bondi, *J. Phys. Chem.* **1964**, *68*, 441-451.

[33] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[34] S. Ma, M. M. J. Smulders, Y. R. Hristova, J. K. Clegg, T. K. Ronson, S. Zarra, J. R. Nitschke, *J. Am. Chem. Soc.* **2013**, *135*, 5678-5684.

[35] S. P. Black, D. M. Wood, F. B. Schwarz, T. K. Ronson, J. J. Holstein, A. R. Stefankiewicz, C. A. Schalley, J. K. M. Sanders, J. R. Nitschke, *Chem. Sci.* **2016**, *7*, 2614-2620.

[36] M. Frank, L. Krause, R. Herbst-Irmer, D. Stalke, G. H. Clever, *Dalton Trans.* **2014**, *43*, 4587-4592.

[37] K. Osowska, O. Š. Miljanić, *J. Am. Chem. Soc.* **2011**, *133*, 724-727.

[38] Y.-R. Zheng, H.-B. Yang, B. H. Northrop, K. Ghosh, P. J. Stang, *Inorg. Chem.* **2008**, *47*, 4706-4711.

[39] A. Wu, L. Isaacs, *J. Am. Chem. Soc.* **2003**, *125*, 4831-4835.

[40] R. Chakrabarty, P. S. Mukherjee, P. J. Stang, *Chem. Rev.* **2011**, *111*, 6810-6918.

[41] T. Weilandt, R. W. Troff, H. Saxell, K. Rissanen, C. A. Schalley, *Inorg. Chem.* **2008**, *47*, 7588-7598.

[42] F. J. Rizzuto, W.-Y. Wu, T. K. Ronson, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2016**, *55*, 7958-7962.

[43] F. J. Rizzuto, J. R. Nitschke, *Nat. Chem.* **2017**, *9*, 903-908.

[44] I. A. Riddell, M. M. J. Smulders, J. K. Clegg, Y. R. Hristova, B. Breiner, J. D. Thoburn, J. R. Nitschke, *Nat. Chem.* **2012**, *4*, 751-756.

[45] R. L. Paul, Z. R. Bell, J. C. Jeffery, J. A. McCleverty, M. D. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4883-4888.

[46] Z. R. Bell, J. C. Jeffery, J. A. McCleverty, M. D. Ward, *Angew. Chem. Int. Ed.* **2002**, *41*, 2515-2518.

[47] P. Dugourd, R. R. Hudgins, D. E. Clemmer, M. F. Jarrold, *Rev. Sci. Instrum.* **1996**, *68*, 1122-1129.

[48] Y.-T. Chan, X. Li, M. Soler, J.-L. Wang, C. Wesdemiotis, G. R. Newkome, *J. Am. Chem. Soc.* **2009**, *131*, 16395-16397.

[49] L. Wang, Z. Zhang, X. Jiang, J. A. Irvin, C. Liu, M. Wang, X. Li, *Inorg. Chem.* **2017**.

[50] E. Kalenius, M. Groessl, K. Rissanen, *Nat. Rev. Chem.* **2018**.

[51] R. A. Bilbeisi, J. K. Clegg, N. Elgrishi, X. de Hatten, M. Devillard, B. Breiner, P. Mal, J. R. Nitschke, *J. Am. Chem. Soc.* **2012**, *134*, 5110-5119.

[52] T. K. Ronson, C. Giri, N. Kodiah Beyeh, A. Minkkinen, F. Topić, J. J. Holstein, K. Rissanen, J. R. Nitschke, *Chem.—Eur. J.* **2013**, *19*, 3374-3382.

[53] W. Meng, T. K. Ronson, J. R. Nitschke, *Proc. Nat. Acad. Sci.* **2013**, *110*, 10531-10535.

[54] J. K. Clegg, J. Cremers, A. J. Hogben, B. Breiner, M. M. J. Smulders, J. D. Thoburn, J. R. Nitschke, *Chem. Sci.* **2013**, *4*, 68-76.

Chapter 7

General conclusions

Metal-organic cages are good candidates for applications like drug delivery thanks to their ability to selectively encapsulate and stabilise guest molecules. In particular, the combination of coordination cages with biomolecules could facilitate their uptake in the body and progress their potential use as drug carriers. This thesis has explored the interactions of self-assembled cages with biomolecules and the properties that resulted from the synergy between these compounds.

Chapter Three focused on the incorporation of tetrahedral coordination cages in supramolecular gels made from tripeptides through non-covalent linkages. A gel matrix containing homogeneously embedded cages was formed when the tripeptide was self-assembled in the presence of the cages. While the cages influenced the gel properties at high concentrations, their ability to encapsulate chemicals was retained, increasing the gel selectivity for guests. By layering gels containing cages with affinities for different guests on top of each other, spatial separation of chemicals was achieved within a single solvent system.

The covalent linkage of biomolecules onto coordination complexes was studied in the following chapters. While attachment of amino acids and peptides directly through their N-termini was unsuccessful, these molecules were tagged to cages via *p*-aminobenzoyl linkers. In Chapter Four, the tripeptide gelator previously used was covalently attached to a tetrahedral cage. The gel formed by the biofunctionalised cage was sensitive to changes in acidity and a reversible sol-gel transition was light-triggered by the introduction of a long-lived photo-acid generator into the system. In Chapter Five, both *N*-acetylgalactosamine and fluorescent moieties were synthesized and conjugated onto a water-soluble cage. The functionalised cage was stable at low concentrations in PBS and its ability to be internalised in hepatocytes through the asialoglycoprotein receptor will be investigated in the future.

As more complex synthetic systems are created from self-assembling molecules in order to mimic their biological counterparts, the necessity for better characterisation and understanding of multicomponent self-sorting is growing. Chapter Six was thus dedicated to the investigation of mass spectrometry as a potential tool for the quantitative analysis of self-sorted systems. The speciation that occurred in three different chemical libraries was studied and rules governing the self-assembly of structures incorporating different types of ligand were drawn. The free Gibbs energies of each congener of the libraries were also obtained for two of the systems studied.